TYPE-k NEUTROSOPHIC SETS

Recursive Triadic Structures, Tensorial Extensions, and Epistemic Auditing of Large Language Models

T / I / F recursion from scalar judgement to auditable epistemic structure

Florentin Smarandache
Maikel Leyva-Vazquez

NSIA Publishing | 2026

Type-k Neutrosophic Sets

Recursive Triadic Structures, Tensorial Extensions, and Epistemic Auditing of Large Language Models

Florentin Smarandache

Department of Mathematics, University of New Mexico, USA

Maikel Leyva-Vazquez

Universidad Bolivariana del Ecuador / Universidad de Guayaquil / Universidad Bernardo O'Higgins

NSIA Publishing

2026

ISBN 978-1-972502-22-8

Table of Contents

Preface

PART I — FOUNDATIONS

Chapter 1. Introduction: The Problem of Nested Uncertainty

Chapter 2. Classical Neutrosophic Sets: A Review

Chapter 3. Type-k Neutrosophic Sets: Recursive Triadic Structures

PART II — TENSORIAL EXTENSIONS

Chapter 4. SVN Tensors: From Vectors to Multilinear Structures

Chapter 5. Plithogenic Tensors: A Hierarchical Multilinear Framework

Chapter 6. Decision Optimality under Epistemic Indeterminacy

PART III — NEUTROSOPHIC PARACONSISTENT LOGIC

Chapter 7. Neutrosophic Paraconsistent Logic: Formal System

Chapter 8. NPL and Type-k: Connections and Alignments

PART IV — EPISTEMIC AUDITING OF LARGE LANGUAGE MODELS

Chapter 9. Breaking the Chains of Probability

Chapter 10. Hallucination Detection Beyond Softmax: The Trichotomy

Chapter 11. A Unified Epistemic Auditing Protocol

PART V — SYNTHESIS AND OPEN PROBLEMS

Chapter 12. Open Problems and Research Directions

References

Preface

This book brings together a programme of work in neutrosophic logic and its tensorial extensions, developed jointly by the authors during the period 2024-2026. The unifying thread is the conviction that contemporary artificial intelligence systems, and large language models in particular, produce responses whose epistemic structure is irreducibly triadic: simultaneously true to some degree, false to some degree, and indeterminate to some degree, with no constraint that these three components sum to unity. Classical probability theory, fuzzy logic, and even intuitionistic fuzzy frameworks force a one-dimensional projection of this triadic structure. The cost of that projection is the collapse of distinctions that operationally matter: between ignorance and paradox, between vagueness and ethical contradiction, between an evaluator who has examined a question and one who refuses to engage.

The book is organised in five parts. Part I lays the formal foundations of Type-k Neutrosophic Sets, a recursive extension in which each epistemic component is itself characterised by a full neutrosophic triplet. Part II develops the tensorial machinery (SVN tensors and plithogenic tensors) that makes the framework computationally tractable for multi-criteria decision-making and large-scale evaluation pipelines. Part III introduces Neutrosophic Paraconsistent Logic, a hybrid formal system that extends da Costa's annotated paraconsistent logic with the neutrosophic indeterminacy component and formalises the distinction between ontological and epistemic contradiction. Part IV applies the framework to the empirical auditing of large language models, presenting both theoretical impossibility results (the softmax-paraconsistency theorem) and large-scale empirical studies that involve more than five thousand evaluations across six foundation models and five elicitation protocols. Part V closes the volume with ten open problems and a research agenda for the period 2026-2030.

The book is designed to be read either linearly or modularly. Readers interested in the formal mathematics may concentrate on Parts I-III. Readers interested in the empirical evaluation of language models may begin with Part IV and consult the formal chapters only as needed. Practitioners of multi-criteria decision-making in industrial engineering will find Chapter 6 self-contained. Philosophers of logic will find Chapters 7 and 8 the most directly relevant to the paraconsistency literature.

We acknowledge institutional support from the University of New Mexico, the Universidad de Guayaquil, the Universidad Bolivariana del Ecuador, and the Universidad Bernardo O'Higgins. We thank Tony Mason for the open release of his data and code on cross-vendor neutrosophic evaluation, which was decisive in motivating the empirical chapters. The errors that remain are entirely our own.

Florentin Smarandache and Maikel Leyva-Vazquez

Gallup, New Mexico and Guayaquil, Ecuador

May 2026

PART I

Foundations

Chapter 1

Introduction: The Problem of Nested Uncertainty

The deployment of large language models in high-stakes decision domains has made the rigorous quantification of epistemic uncertainty a first-order engineering requirement. A clinical decision support system may simultaneously exhibit high diagnostic accuracy on cases similar to those in its training distribution, genuine unknown failure modes on cases that are out of distribution, and a documented hallucination risk that varies non-monotonically with prompt structure. A judicial reasoning system may simultaneously cite a binding precedent, acknowledge a genuine conflict of legal principles, and confess that the controlling jurisprudence is unsettled. Collapsing these three independent dimensions into a single scalar score, as the dominant probabilistic frameworks of contemporary machine learning require, destroys exactly the information that downstream decision makers need.

1.1 The Problem

Contemporary uncertainty quantification for artificial intelligence systems rests almost entirely on probability theory. A model is asked to produce, for each candidate answer to a question, a number in the unit interval that represents the probability of that answer being correct. The probabilities are required to sum to unity across the set of candidate answers, a constraint that the softmax normalisation enforces architecturally at the output layer of almost every modern neural classifier. Within this paradigm, an increase in the probability assigned to one answer must be exactly compensated by a decrease in the probability assigned to others. There is no room for an evaluator to say that a particular answer is highly supported, that its negation is also highly supported, and that the question itself is structurally indeterminate. The probabilistic framework forces the evaluator to choose.

This forced choice has empirical consequences that we document throughout this book. When models are released from the softmax constraint and asked to provide three independent assessments, one for truth, one for indeterminacy, and one for falsity, with no requirement that the three values sum to anything in particular, the responses they produce are systematically different from those produced under the probabilistic protocol. Across the empirical studies reported in Parts IV of this book, between 24.7 percent and 84 percent of responses (depending

on the protocol) violate the sum-to-unity constraint, and they do so in ways that are concentrated precisely on the cases where the probabilistic framework is least informative: ethical contradictions, paradoxes of self-reference, statements about future contingents, and cases of genuine epistemic ignorance.

A second, deeper problem also arises. Once an evaluator has produced a triplet (T, I, F) of independent assessments, a natural question follows immediately: to what degree is T itself true, indeterminate, or false? The evaluator has assessed the truth of the original proposition, but it has not assessed the reliability of its own assessment. In type-2 fuzzy logic, this second-order question is handled by allowing membership degrees themselves to be fuzzy sets. The neutrosophic analogue, in which each component of the triplet (T, I, F) is itself characterised by a full neutrosophic triplet, had not been formally defined in the literature prior to the work we present in this book. Type-k Neutrosophic Sets, the central formal object of Part I, fill that gap.

1.2 Why Classical Neutrosophy is Insufficient

Smarandache's classical formulation of neutrosophic logic, introduced in 1998, already addresses the main limitation of probability theory by allowing T, I, and F to vary independently in the unit interval. This is a substantial advance over both fuzzy logic, which conflates falsity with the complement of truth, and intuitionistic fuzzy logic, which constrains the sum T + F to be at most unity and forces the residual to be a dependent hesitation degree. Single-Valued Neutrosophic Sets (SVNS), introduced by Wang et al. in 2010, provide the standard scalar realisation in which T, I, F are all real numbers in $[0, 1]$ without any sum constraint.

Three limitations of the classical neutrosophic framework motivate the extensions developed in this book. First, the SVNS framework treats each component as a scalar; it provides no mechanism to represent the evaluator's uncertainty about its own assessment. Second, classical neutrosophic aggregation operators treat the criteria of a multi-criteria evaluation as independent; they offer no mechanism for representing that two criteria are related, redundant, or in structural contradiction. Third, the framework restricts each component to the unit interval $[0, 1]$. Empirical observation of language model behaviour under extended elicitation protocols reveals that the models, when permitted, produce responses with components outside this range: indeterminacy values above one, truth values above one, even occasionally negative truth values that the standard formulation cannot accommodate.

Each of these three limitations is addressed by a separate extension developed in the chapters that follow. The recursive Type-k construction of Chapter 3 addresses the first limitation by allowing each component to itself be characterised by a neutrosophic triplet at the next level of the recursion. The plithogenic tensor framework of Chapters 4 and 5 addresses the second limitation by introducing an explicit contradiction function on pairs of criteria, so that aggregation respects the structural tensions among them. The overset, underset, and offset regimes introduced by Smarandache in 2016 address the third limitation by extending the admissible range of T, I, F to intervals such as $[0,2]$ or $[-1,2]$, and Chapters 9 through 11 document the empirical realisability of these extended ranges in current foundation models.

1.3 Roadmap of the Book

The book is organised as follows. Part I, comprising the present chapter and Chapters 2 and 3, lays the formal foundations. Chapter 2 reviews the classical theory of neutrosophic sets, the single-valued specialisation, and their established applications in multi-criteria decision making. Chapter 3 introduces Type-k Neutrosophic Sets, proves the strict expressive hierarchy theorem, and presents empirical evidence that, when evaluation protocols permit components outside the unit interval, current language models routinely produce them — a necessary condition for Type-2 representation rather than direct proof of second-order computation.

Part II develops the tensorial extensions. Chapter 4 introduces single-valued neutrosophic tensors, the multilinear-algebraic generalisation that supports Tucker decomposition, cut-tensor representation, and Einstein contraction. Chapter 5 enriches these objects with a contradiction function over pairs of attributes, yielding plithogenic tensors and the formal foundation for the response to recent critiques of declared-loss evaluation. Chapter 6 develops the decision-theoretic consequences: the Decision Optimality theorem, which shows that SVN tensors can reverse intuitionistic-fuzzy rankings when indeterminacy is non-zero, and the Contradiction Visibility theorem, which shows that plithogenic tensors expose inter-criterion tensions that change recommendations in safety-critical scenarios.

Part III, comprising Chapters 7 and 8, introduces Neutrosophic Paraconsistent Logic. Chapter 7 develops the formal system, proves the eight central theorems including the strict-extension results over both da Costa's annotated paraconsistent logic and Smarandache's single-valued neutrosophic logic, and validates the formalism on three domains: indigenous rights in Ecuador,

cross-cultural AI alignment, and the wave-particle duality of quantum mechanics. Chapter 8 establishes the connection between Neutrosophic Paraconsistent Logic and Type-k Neutrosophic Sets, mapping the seven epistemic states of the former to specific configurations in the latter.

Part IV turns to the empirical auditing of large language models. Chapter 9 reports a study of three hundred API calls across four OpenAI models, demonstrating hyper-truth ($T + I + F > 1(hyper - truth\ region \mathcal{H} \subset [0,1]^3)$) in 66 percent of evaluations under the neutrosophic strategy and zero percent under the probabilistic strategy. Chapter 10 proves the softmax-paraconsistency impossibility theorem: any hallucination detector that derives both its truth and its falsity signals from the same softmax-normalised natural-language inference head is structurally incapable of detecting responses in the simultaneous-evidence regime. A dual-NLI protocol that recovers the excluded regime is presented and validated on a fifty-pair synthetic benchmark. Chapter 11 synthesises the protocols of Chapters 9 and 10 with the Type-k framework of Chapter 3 to propose a unified epistemic auditing pipeline.

Part V, comprising the closing Chapter 12, lists ten open problems and outlines a research agenda for the period 2026-2030. The agenda connects the formal work of Parts I-III with the empirical programme of Part IV, and situates both in the broader context of pluriversal AI alignment, neutrosophic statistics, and the emerging field of plithogenic evaluation for high-stakes decision support.

1.4 Conclusion

The central thesis of this book is that the epistemic structure of artificial intelligence systems is irreducibly triadic and recursive: triadic because truth, indeterminacy, and falsity are independent dimensions that cannot be reduced to a single scalar, and recursive because the evaluator's assessment of each dimension is itself subject to evaluation at the next level of the analysis. The classical neutrosophic framework provides the triadic apparatus; the Type-k extension introduced in Chapter 3 provides the recursive apparatus; the tensorial extensions of Part II provide the computational machinery; the paraconsistent extensions of Part III provide the inferential machinery; and the empirical chapters of Part IV demonstrate that all of this apparatus is operationally necessary for the auditing of contemporary foundation models. The remainder of the book develops each component in the detail that the working researcher and practitioner will need.

Chapter 2

Classical Neutrosophic Sets: A Review

Before introducing the Type-k extension that is the central contribution of Part I, we review the classical theory of neutrosophic sets, the single-valued specialisation that is most directly relevant to applications, and the standard aggregation operators that have become canonical in the literature on multi-criteria decision-making. The exposition follows Smarandache (1998, 2005, 2010) and Wang et al. (2010) with notation adapted for compatibility with the tensor extensions developed in subsequent chapters.

2.1 Definition of Neutrosophic Sets

Neutrosophy, introduced by Smarandache in 1998, generalises both fuzzy logic and intuitionistic fuzzy logic by representing each proposition or membership relation as an ordered triplet of independent components.

> **Definition 2.1 (Neutrosophic Set).** *Let U be a universe of discourse. A neutrosophic set A on U assigns to each element x in U an ordered triplet (T(x), I(x), F(x)) where T(x) is the truth-membership degree, I(x) is the indeterminacy-membership degree, and F(x) is the falsity-membership degree. In the most general formulation, T(x), I(x), and F(x) are subsets of the non-standard unit interval, and no constraint is imposed on their sum.*

The three components are required to be independent in the sense that no logical or arithmetic constraint links them. This is the fundamental departure from intuitionistic fuzzy logic, in which the truth and falsity components are constrained to sum to at most unity and the hesitation degree is forced to be the residual. In a neutrosophic set, an element may have truth degree 0.9, falsity degree 0.8, and indeterminacy degree 0.7 simultaneously: the three components together describe an epistemic state in which the evidence both for and against the proposition is strong and the residual uncertainty is also substantial. No constraint requires that they sum to anything in particular.

2.2 Single-Valued Neutrosophic Sets

For computational and applied purposes, the most widely used specialisation is the single-valued neutrosophic set, introduced by Wang, Smarandache, Zhang, and Sunderraman in 2010.

Definition 2.2 (Single-Valued Neutrosophic Set, SVNS). *A single-valued neutrosophic set A on a universe U is the set of ordered quadruples A = { (x,* $T_A(x)$*,* $I_A(x)$*,* $F_A(x)$*) : x in U }, where* $T_A(x)$*,* $I_A(x)$*,* $F_A(x)$ *are real numbers in the closed unit interval* $[0, 1]$ *and no constraint is imposed on their sum, which therefore lies in the interval [0, 3].*

The single-valued specialisation replaces the non-standard subsets of the general definition with ordinary real numbers in the unit interval, sacrificing some expressive generality in exchange for computational tractability. Every operation defined on general neutrosophic sets has a single-valued specialisation; throughout this book, except where otherwise noted, we work with single-valued neutrosophic sets and refer to them simply as neutrosophic sets where context makes the specialisation unambiguous.

2.3 Basic Properties and Operations

The standard operations on single-valued neutrosophic sets are componentwise extensions of the corresponding fuzzy operations, using the t-conorm-product and t-norm-product operators.

Definition 2.3 (SVNS operations). *For SVNS elements* $a = (T_1, I_1, F_1)$ *and* $b = (T_2, I_2, F_2)$*: (i) addition:* $a + b = (T_1 + T_2 - T_1 T_2, I_1 I_2, F_1 F_2)$*; (ii) multiplication:* $a \cdot b = (T_1 T_2, I_1 + I_2 - I_1 I_2, F_1 + F_2 - F_1 F_2)$*; (iii) scalar product:* $\lambda \cdot a = (1 - (1 - T_1)^\lambda, I_1^\lambda, F_1^\lambda)$*; (iv) complement:* $a^c = (F_1, 1 - I_1, T_1)$*.*

These operations satisfy the algebraic properties of commutativity, associativity, and idempotence in the appropriate senses, and they reduce to the standard fuzzy operations when I = 0 and F = 1 - T. The complement operation deserves particular attention: in SVNS, the complement of a triplet swaps the truth and falsity components and reflects the indeterminacy around its midpoint. This formulation preserves the independence of indeterminacy from truth-falsity, in contrast to fuzzy and intuitionistic fuzzy frameworks where the complement is constrained by the underlying sum relation.

2.4 The Score, Accuracy, and Certainty Functions

For applications that require a total ordering on neutrosophic elements, several score functions have been proposed in the literature. The most widely used is the function introduced by Ye in 2014.

Definition 2.4 (Score, accuracy, and certainty functions). *The score function (Ye 2014) of a single-valued neutrosophic element* $a = (T, I, F)$ *is:*

$$S(a) = \frac{2 + T - I - F}{3}, \; S(a) \in [0, 1].$$

The score function alone induces only a pre-order, since two distinct triplets may receive the same score. To obtain a total order on the set of neutrosophic triplets — indispensable whenever a multi-criteria decision must return an unambiguous ranking of alternatives — Smarandache (2025) complements the score with two further functions, applied in cascade. The accuracy function of a single-valued neutrosophic element $a = (T, I, F)$ is:

$$A(a) = T - F, \;\; A(a) \in [-1,1].$$

and the certainty function is:

$$C(a) = T, \;\; C(a) \in [0,1].$$

The accuracy function measures the net balance between truth and falsity, disregarding the indeterminacy, whereas the certainty function measures the degree of truth-membership alone. Together with the score they resolve ties in a fixed lexicographic priority.

Proposition 2.1 (Total order on neutrosophic triplets, Smarandache 2025). The score, accuracy, and certainty functions, applied lexicographically in this order, determine a total order on the set of neutrosophic triplets (T, I, F). For two single-valued neutrosophic elements $a_1 = (T_1, I_1, F_1)$ and $a_2 = (T_2, I_2, F_2)$:

(i) $S(a_1) > S(a_2) \Rightarrow a_1 > a_2$;

(ii) $S(a_1) = S(a_2) \;\wedge\; A(a_1) > A(a_2) \Rightarrow a_1 > a_2$;

(iii) $S(a_1) = S(a_2) \;\wedge\; A(a_1) = A(a_2) \;\wedge\; C(a_1) > C(a_2) \Rightarrow a_1 > a_2$;

(iv) $S(a_1) = S(a_2) \;\wedge\; A(a_1) = A(a_2) \;\wedge\; C(a_1) = C(a_2) \Rightarrow a_1 = a_2$.

The construction extends componentwise to interval-valued and, more generally, subset-valued neutrosophic triplets, and supplies the tie-breaking discriminator used in multi-criteria decision-making (MCDM) when the score function alone yields equal values.

The score function provides a deterministic ranking of SVNS elements that respects the intuition that higher truth and lower indeterminacy and falsity correspond to higher score. It is the standard tool for converting the three-dimensional epistemic state into a one-dimensional ranking when a decision must be made. The cost of the conversion is the loss of information that the three independent components carry; the Plithogenic Truth Score introduced in Chapter 6 partially recovers this lost information by additionally penalising inter-criterion contradiction.

2.5 The SVNWA Aggregation Operator

Multi-criteria decision-making requires aggregating multiple neutrosophic assessments, typically one per criterion, into a single overall assessment per alternative. The canonical aggregation operator is the Single-Valued Neutrosophic Weighted Average (SVNWA), introduced by Ye (2014) and adopted as the standard in the subsequent literature.

Definition 2.5 (SVNWA aggregation operator). Given n single-valued neutrosophic numbers $a_i = (T_i, I_i, F_i)$ for $i = 1, \dots, n$ with non-negative weights w_i satisfying $\sum_{i=1}^{n} \quad w_i = 1$, the Single-Valued Neutrosophic Weighted Average is

$$\mathrm{SVNWA}(a_1, \dots, a_n) = (1 - \prod_{i=1}^{n}(1 - T_i)^{w_i}, \prod_{i=1}^{n} I_i^{w_i}, \prod_{i=1}^{n} F_i^{w_i}).$$

The SVNWA operator satisfies the standard axioms for an aggregation operator: idempotence (when all inputs are equal, the output equals the common value), monotonicity (componentwise), and boundedness (the output is between the componentwise minimum and maximum of the inputs). It reduces to the standard weighted geometric mean on the indeterminacy and falsity components and to a t-conorm-product average on the truth component. The recursive Type-k extension presented in Chapter 3 generalises SVNWA to operate on nested triplets while preserving backward compatibility with the standard operator as a degenerate case.

2.6 Applications in Multi-Criteria Decision-Making

Single-valued neutrosophic sets have been applied extensively in multi-criteria decision-making over the past decade. Abdel-Basset and collaborators developed a neutrosophic adaptation of the TOPSIS method for supplier selection in supply chain management and demonstrated superior discrimination power compared with interval-valued and intuitionistic fuzzy approaches. Biswas,

Pramanik, and Giri developed a TOPSIS variant for multi-attribute group decision-making under the single-valued neutrosophic environment. Ye introduced simplified neutrosophic weighted aggregation operators that have become standard in engineering decision problems. The hybrid plithogenic approach with quality function deployment for supply chain sustainability metrics, developed by Abdel-Basset and collaborators, foreshadows the plithogenic extensions developed in Part II of this book.

These applications share a common structure. A decision-maker is presented with m alternatives to be evaluated against n criteria. Each (alternative, criterion) pair is assessed by one or more experts who provide a single-valued neutrosophic triplet capturing the truth, indeterminacy, and falsity of the assertion that the alternative satisfies the criterion. The triplets are aggregated across experts (typically by a weighted operator that reflects expert credibility) and then across criteria (by SVNWA or a related operator that reflects criterion importance). The aggregated triplet for each alternative is converted to a scalar by the score function, and the alternatives are ranked. The framework has been validated across domains as diverse as smart medical device selection, location decisions, sustainability assessment, and reliability engineering.

2.7 Conclusion

Classical single-valued neutrosophic sets provide a triadic framework in which truth, indeterminacy, and falsity are independent dimensions in the unit interval. The associated algebraic operations, score function, and SVNWA aggregation operator constitute a coherent and tested machinery for multi-criteria decision-making. Three limitations of this framework, however, motivate the extensions developed in the remainder of this book. First, the framework treats each component as a scalar and provides no mechanism for representing the evaluator's uncertainty about its own assessment; this is the limitation that the recursive Type-k construction of Chapter 3 addresses. Second, the framework treats criteria as independent and provides no mechanism for representing inter-criterion contradiction; this is the limitation that the plithogenic tensor framework of Chapters 4 and 5 addresses. Third, the framework restricts each component to the unit interval and provides no mechanism for representing the extended regimes that empirical evaluation of language models reveals to be operationally necessary; this is the limitation that the overset, underset, and offset protocols of Chapter 9 address.

Chapter 3

Type-k Neutrosophic Sets: Recursive Triadic Structures

This chapter introduces the central formal object of Part I: the Type-k Neutrosophic Set. We define the recursive construction, prove the strict expressive hierarchy theorem, develop the canonical embedding that ensures backward compatibility with Type-1 (classical) neutrosophic sets, derive the recursive Single-Valued Neutrosophic Weighted Average operator, and present empirical evidence from 2,419 evaluations of six foundation models that, when the elicitation protocol permits values outside the unit interval, current large language models routinely produce them — a necessary condition for Type-2 representation, not by itself proof of genuine second-order computation. The chapter closes with an analysis of the implications for industrial multi-criteria decision-making systems that incorporate language-model expert opinions.

3.1 Motivation and Recursive Definitions

The classical neutrosophic framework of Chapter 2 assigns to each element a triplet (T, I, F) of scalar components. A natural question arises immediately: to what degree is T itself true, indeterminate, or false? The evaluator has assessed the truth of the underlying proposition, but it has not assessed the reliability of its own assessment. In type-2 fuzzy logic, introduced by Mendel and John in 2002, this second-order question is handled by allowing membership degrees themselves to be fuzzy sets. The neutrosophic analogue, in which each component of the triplet (T, I, F) is itself characterised by a full neutrosophic triplet, had not been formally defined in the literature prior to the present work. We now provide that definition.

Definition 3.1 (Type-1 Neutrosophic Set). A Type-1 Neutrosophic Set on a universe U assigns to each element x in U a triplet $\mu^{(1)}(x) = (T, I, F)$,T, I, F $\in$ [0,1] independently and without sum constraint. $3^1 = 3$ scalars. Equivalently, $NS^{(1)}(U)$ is the standard single-valued neutrosophic set introduced in Definition 2.2.

Definition 3.2 (Type-2 Neutrosophic Set). *A Type-2 Neutrosophic Set $A^{(2)}$ on U assigns to each x in U nine scalars organised as three nested triplets. The truth component T is itself characterised by a triplet (T_T, I_T, F_T), where T_T is the truth of T, I_T is the indeterminacy of T,*

and F_T is the falsity of T. The indeterminacy component I is characterised by (T_I, I_I, F_I), and the falsity component F is characterised by (T_F, I_F, F_F). All nine scalars take values in $[0, 1]$ *independently.* 3^2 *= 9 scalars.*

Definition 3.3 (Type-3 Neutrosophic Set). *Each of the nine scalars of a Type-2 representation is itself replaced by a neutrosophic triplet, yielding* 3^3 *= 27 scalars per element. For instance, the sub-component T_T is replaced by (T_{TT}, I_{TT}, F_{TT}), and analogously for all sub-components.*

Definition 3.4 (Type-k Neutrosophic Set). *A Type-k Neutrosophic Set is defined recursively. The base case k = 1 is the standard neutrosophic set of Definition 3.1. For k at least 2, each scalar component of a Type-(k-1) element is replaced by a full neutrosophic triplet. Each element x in U is therefore characterised by* 3^k *scalars. The collection of Type-k neutrosophic sets over U is denoted $NS^{(k)}$(U).*

The construction admits a particularly compact alternative formulation, due to Smarandache, that captures the combinatorial structure of the recursion in a single line. We state it as a remark.

Remark 3.1 (Smarandache's combinatorial formulation). *A Type-k Neutrosophic Set, denoted $\mathcal{NS}^{(k)}$, is the set of all strings $c_1 c_2 \dots c_k$ formed by concatenating k symbols, where each $c_i \in T, I, F$. Hence the set contains exactly* 3^k *elements:* $\left|\mathcal{NS}^{((k))}\right| = 3^k$

The combinatorial formulation makes the structure manifest. At depth k, each scalar component of the representation is labelled by a string of length k over the three-letter alphabet {T, I, F}. The string $c_1 c_2 \dots c_k$ labels the scalar that captures the c_k-aspect of the c_{k-1}-aspect of the ... of the c_1-aspect of the original proposition. The total number of such strings is 3^k, exactly matching the cardinality stated in Definition 3.4.

Remark 3.2 (Infinite-dimensional case). *More generally, T, I, F may be neutrosophic sets rather than scalars, yielding an infinite-dimensional construction that subsumes all finite Type-k as projections. The formal development of this case, including the convergence conditions for recursive aggregation, is left for future work.*

3.2 Canonical Embedding and Strict Hierarchy

For the Type-k construction to be useful in practice, two structural properties are required. First, every Type-(k-1) element must embed canonically into the Type-k space, so that existing Type-1 models and data can be lifted to Type-k without information loss. Second, the embedding must be strict: Type-k must be more expressive than Type-(k-1), so that the lift actually adds something. We establish both properties in this section.

Definition 3.5 (Canonical embedding φ). *The canonical embedding $\phi: \mathcal{NS}^{((k-1))} \hookrightarrow \mathcal{NS}^{((k))}$ maps each scalar component s to the triplet $\phi(s) = (s, 0, 1 - s)$. Under this embedding, a Type-1 element (T, I, F) maps to the Type-2 element with sub-triplets ((T, 0, 1 - T), (I, 0, 1 - I), (F, 0, 1 - F)), expressing full certainty about each component.*

The canonical embedding captures a precise intuition: when we lift a Type-(k-1) element to Type-k, we assert that we are fully certain about each component of the original representation. The truth component T is asserted with truth s = T (we are as confident as the value itself indicates), indeterminacy 0 (no second-order doubt), and falsity 1 - T (the second-order falsity is the complement of the value). This is the most informationally conservative lift available, and it is the lift that recovers the standard Type-1 behaviour as a degenerate case of Type-k.

Theorem 3.1 (Strict hierarchy). *The inclusions $NS^{(1)}(U) \subsetneq NS^{(2)}(U) \subsetneq \cdots \subsetneq NS^{(k)}(U)$ hold for every finite k. That is, Type-k is strictly more expressive than Type-(k-1) for every k at least 2.*

Proof. The inclusion follows from the canonical embedding φ of Definition 3.5: every Type-(k−1) element has a well-defined image in Type-k, so $NS^{(k-1)}(U) \subseteq NS^{(k)}(U)$. For strictness at $k = 2$ it suffices to show that φ is not surjective. Its image consists exactly of those Type-2 elements whose three sub-triplets all have the canonical form $(s, 0, 1 - s)$ — that is, with zero second-order indeterminacy and second-order falsity equal to the complement of second-order truth. Now consider the Type-2 element whose truth component is characterised by the sub-triplet (0.6, 0.5, 0.3), so that its second-order indeterminacy is $I_T = 0.5 \neq 0$. All nine scalars lie in [0,1], so this is a legitimate Type-2 object; yet it has no Type-1 pre-image under φ, because the embedding forces the second-order indeterminacy of every component to vanish. Hence $NS^{(1)}(U) \subsetneq NS^{(2)}(U)$. Applying the same argument at

the deepest level of the recursion gives $NS^{(k-1)}(U) \subsetneq NS^{(k)}(U)$ for every $k \geq 2$ by induction. ■

3.3 Computational Properties and Tractability

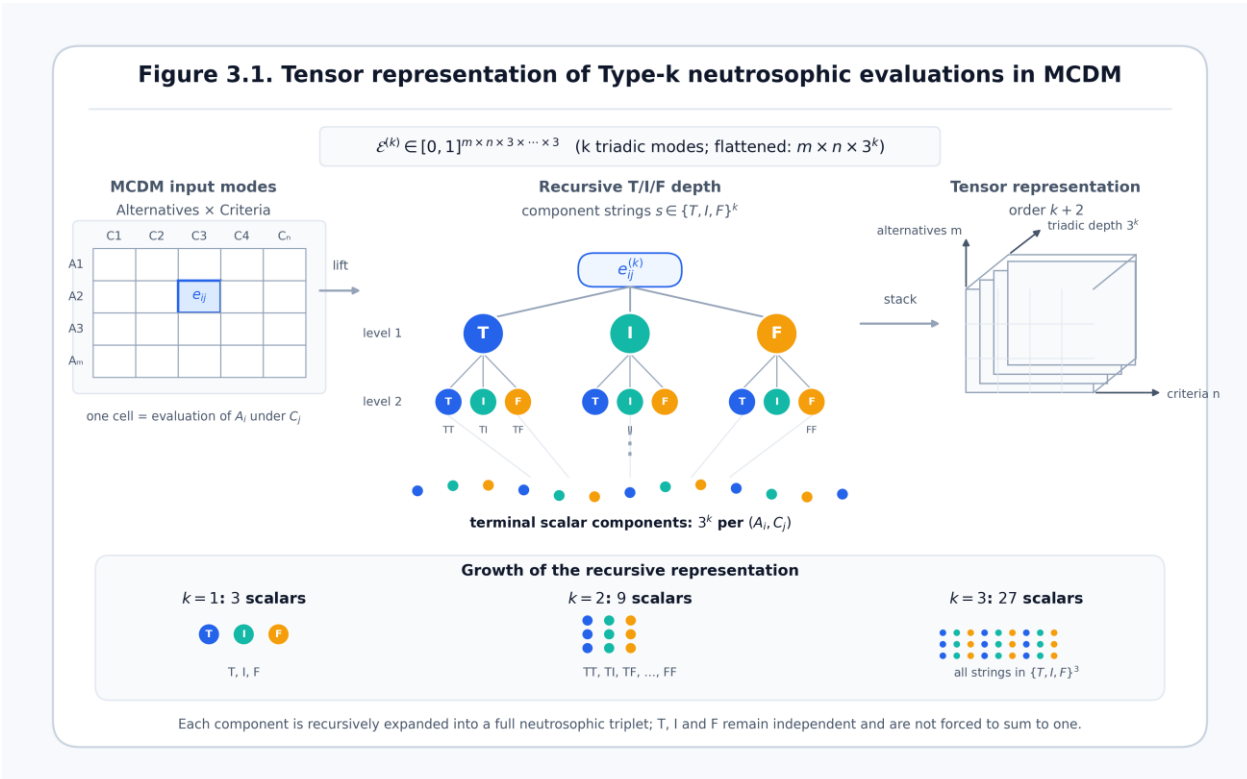

Figure 3.1. Tensor representation of Type-k neutrosophic evaluations in multi-criteria decision-making.

The computational cost of Type-k grows as 3^k scalars per element, which is tractable for the values of k relevant in practice. For k = 1 (standard SVNS) the cost is 3 scalars per element. For k = 2 the cost is 9 scalars, and for k = 3 it is 27 scalars. In industrial multi-criteria decision-making applications, k = 2 is sufficient to capture meta-uncertainty about expert opinion reliability, which is the most common source of nested uncertainty in engineering decision problems, while remaining computationally efficient for problems with up to hundreds of alternatives and criteria.

The tensor structure of Type-k Neutrosophic Sets is aligned with established tensor decomposition methods, as developed systematically in Chapter 4. A Type-k neutrosophic evaluation matrix of m alternatives by n criteria can be represented as a tensor of order k + 2 with dimensions m x n x 3^k, enabling the application of Tucker decomposition and Higher-Order Singular Value Decomposition for dimensionality reduction in large-scale problems. This

connection to tensor algebra opens a rich research agenda for scalable neutrosophic decision methods that we develop further in Part II of this book.

3.4 The Recursive SVNWA Aggregation Operator

For Type-k Neutrosophic Sets to support multi-criteria decision-making, we require an aggregation operator that generalises the standard SVNWA of Definition 2.5 while preserving backward compatibility. We construct such an operator by applying SVNWA recursively and independently to each sub-triplet at every recursion level.

Definition 3.6 (Type-k SVNWA). *The Type-k Single-Valued Neutrosophic Weighted Average is defined recursively. At depth k = 1, $SVNWA^{(1)}$ is the standard operator of Definition 2.5. At depth k at least 2, $SVNWA^{(k)}$ aggregates the three pairs of sub-triplets at the top level of the recursion (the T-sub-triplet, the I-sub-triplet, and the F-sub-triplet) separately and independently using $SVNWA^{(k-1)}$ with the same weight vector w. This preserves the independence of the sub-triplets across the aggregation operation.*

Proposition 3.1 (Backward compatibility). *The Type-k SVNWA reduces exactly to the standard Type-1 SVNWA when all input elements are images under the canonical embedding φ, that is, when all sub-triplets satisfy I = 0 and F = 1 - T at every level beyond the first.*

Proof. By induction on k. The base case k = 1 is trivial. For the inductive step, assume the proposition holds at depth k − 1 and consider Type-k inputs that are images of Type-1 elements under φ. Each Type-k input has the form (φ(T), φ(I), φ(F)) = ((T, 0, 1 − T), (I, 0, 1 − I), (F, 0, 1 − F)). Applying the Type-k SVNWA aggregates the three pairs of sub-triplets independently. The T-sub-triplet (T, 0, 1 − T) is itself under the canonical embedding, so by the inductive hypothesis its aggregation reduces to the standard SVNWA on the T-components alone, yielding $(1 - \prod_i (1 - T_i)^{w_i}, 0, \prod_i (1 - T_i)^{w_i})$. Analogous reductions hold for the I- and F-sub-triplets. The composite output is precisely the canonical embedding of the standard Type-1 SVNWA output, establishing the claim. ■

Proposition 3.1 has a critical practical consequence. Any existing Type-1 multi-criteria decision-making system can be extended to Type-k without modifying the aggregation logic: only the data representation changes, and on Type-1 data the system continues to produce the same results as

before. The lift to Type-k is therefore monotone in information content: it adds new expressive power without breaking any existing application.

The score function for comparing Type-k elements follows naturally from the recursive structure. For Type-2, the score $S^{(2)}$ of an element with sub-triplets is computed by first applying the standard score function $S(T, I, F)$= (2 + T - I - F) / 3 to each sub-triplet, and then applying the standard SVNWA to the three resulting scalars with weights reflecting the relative importance of the meta-uncertainty components. For k > 2, this procedure is applied recursively until the depth reaches the base case.

3.5 Empirical Evidence: When Permitted, LLMs Produce Extended-Range Values

To assess whether the Type-2 extension is empirically necessary or merely a theoretical curiosity, we conducted a systematic multi-vendor evaluation of six state-of-the-art foundation models under five elicitation protocols. The experiment, reported in detail by Smarandache and Leyva-Vazquez (2026), comprised 2,419 usable evaluations and tested whether large language models, when permitted, spontaneously produce responses whose components fall outside the Type-1 admissible range of $[0, 1]$.

3.5.1 Experimental Design

Six foundation models were evaluated: Alibaba Qwen-3-235B, Anthropic Claude Sonnet 4, DeepSeek Chat, Meta Llama-4-Maverick, Mistral Medium-3.1, and OpenAI GPT-4o. These models represent the major commercial AI providers in 2026 and cover a range of architectural families, training regimes, and geopolitical origins, enabling cross-vendor generalisability assessment. Five elicitation protocols were applied to five phenomena (paradox, epistemic ignorance, vagueness, ethical contradiction, and future contingency), plus three tautology controls.

The five protocols are S1 (classical scalar, T, I, F in $[0, 1]$); S4 (Mason's three-component declared-loss frame with explicit loss elicitation, T, I, F in $[0, 1]$); S4-N (per-attribute neutrosophic decomposition); S4-O.A (overset extended range, T, I, F in $[0,2]$); and S4-O.C (offset extended range, T, I, F in $[-1,2]$). A response was classified as requiring Type-2 representation if any component fell outside $[0, 1]$. Each combination was evaluated in triplicate

(three runs per model per protocol per phenomenon), yielding a total of 2,419 usable evaluations. Data and replication code are publicly available at github.com/mleyvaz/typek-neutrosophic-sets.

3.5.2 Results by Protocol

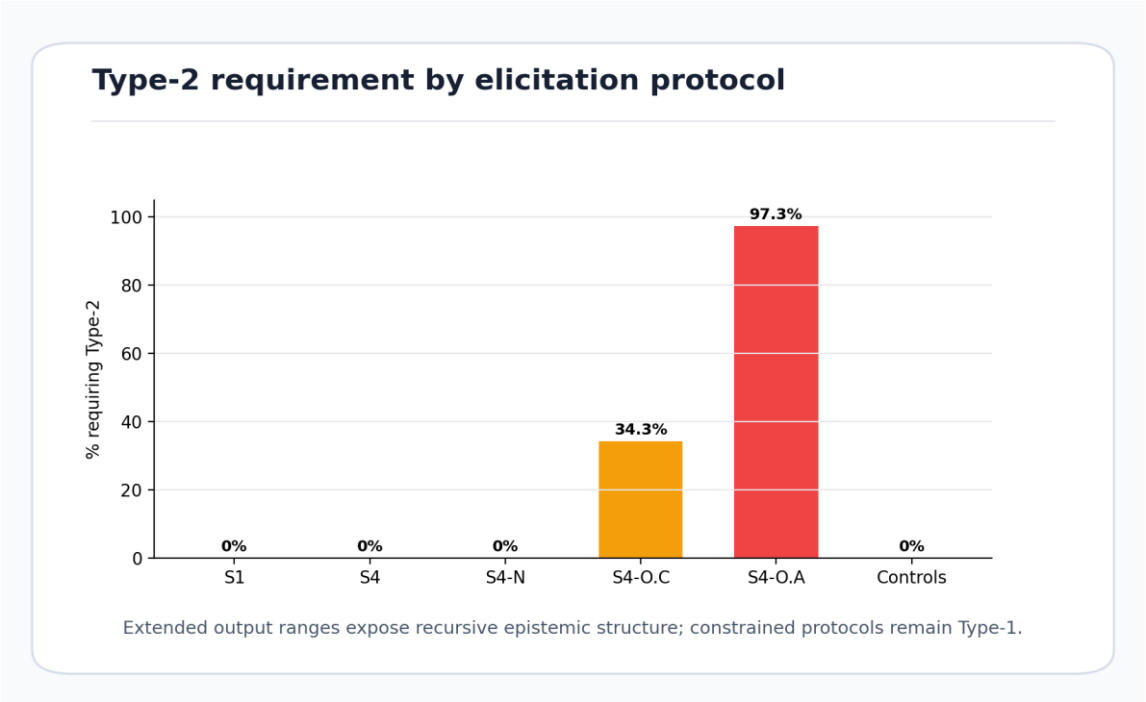

Figure 3.2. Type-2 requirement by elicitation protocol.

Table 3.1 reports the Type-2 requirement rate by elicitation protocol. The pattern is decisive: every $[0, 1]$-constrained protocol (S1, S4, S4-N, and tautology controls) produced zero responses requiring Type-2 representation, while every extended-range protocol produced substantial Type-2 rates. The S4-O.C protocol, with admissible range $[-1,2]$, yielded Type-2 in 34.3 percent of evaluations. The S4-O.A protocol, with admissible range $[0,2]$, yielded Type-2 in 97.3 percent of evaluations. Aggregated across all protocols, 24.7 percent of all responses require Type-2 representation.

Protocol	**Range**	**n**	**Type-2 required**	**%**
S1	[0, 1]	300	0	0.0
S4	[0, 1]	300	0	0.0

S4-N	[0, 1]	300	0	0.0
S4-O.C	$[-1,2]$	889	305	34.3
S4-O.A	$[0,2]$	300	292	97.3
S1-taut (controls)	[0, 1]	90	0	0.0
Total	—	2,419	597	24.7

Table 3.1. Type-2 requirement rate by elicitation protocol.

The complete absence of Type-2 responses under all $[0, 1]$-constrained protocols confirms that Type-2 emergence is a function of the protocol output range, not of prompt wording, model architecture, or the nature of the phenomenon evaluated. When the protocol restricts the admissible range to $[0, 1]$, the models produce only Type-1-compatible responses, with no exceptions across 990 evaluations. When the protocol extends the range, the models begin to produce responses that require the Type-2 framework to represent without loss.

3.5.3 Results by Vendor

Table 3.2 reports the Type-2 requirement rate by vendor on the extended-range protocols (S4-O.A and S4-O.C combined). The range across six architecturally independent vendors is 19.6 to 29.9 percent, confirming a systematic structural phenomenon rather than a single-model artefact.

Vendor	**Model**	**Evaluations**	**Type-2 required**	**%**
Mistral	Medium-3.1	435	130	29.9
OpenAI	GPT-4o	377	100	26.5
Anthropic	Claude Sonnet 4	408	108	26.5
Meta	Llama-4-Maverick	355	81	22.8
DeepSeek	Chat	431	97	22.5
Alibaba	Qwen-3-235B	413	81	19.6

Table 3.2. Type-2 requirement rate by vendor on extended-range protocols.

3.5.4 Distribution of Type-2 Subtypes

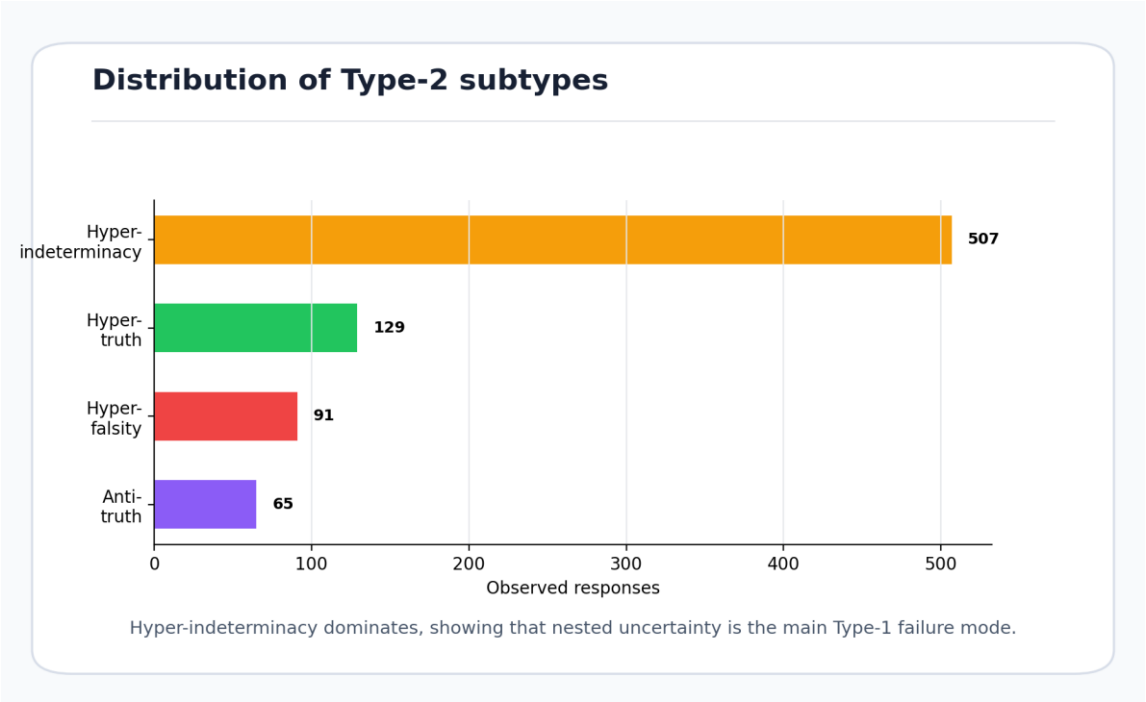

Figure 3.3. Distribution of Type-2 subtypes and their operational interpretation.

The Type-2 responses can be classified into four subtypes by which component falls outside $[0, 1]$ and in which direction. Table 3.3 reports their distribution across the 2,419 evaluations. Hyper-indeterminacy ($I > 1$) is the most frequent subtype at 21.0 percent, indicating that extended indeterminacy is the primary epistemic state that Type-1 cannot represent. Hyper-truth ($T > 1$) and hyper-falsity ($F > 1$) follow at 5.3 and 3.8 percent. Anti-truth ($T < 0$) is the rarest subtype at 2.7 percent but the most striking, since it corresponds to a regime in which the model reports negative truth: the proposition actively subtracts trust rather than supporting it.

Subtype	**Condition**	**Interpretation**	**n**	**% of total**
Hyper-indeterminacy	$I > 1$	Structural uncertainty	507	21.0
Hyper-truth	$T > 1$	Overtime worker regime	129	5.3

Hyper-falsity	F > 1	Paradox under S4-O.A	91	3.8
Anti-truth	T < 0	Adversarial state	65	2.7

Table 3.3. Distribution and operational interpretation of Type-2 subtypes.

3.6 Implications for LLM-Augmented MCDM Systems

The empirical results of Section 3.5 have direct operational consequences for industrial engineering decision systems that use language-model-elicited expert opinions. Consider a standard neutrosophic multi-criteria decision-making workflow: a language model is prompted to evaluate m alternatives across n criteria, returning (T, I, F) triplets for each (alternative, criterion) pair, which are then aggregated using SVNWA and ranked. If the elicitation protocol is S1 or S4 (constrained to $[0, 1]$), the results of Section 3.5 show that 0 percent of responses will require Type-2; the Type-1 workflow is sufficient and the standard tools apply without modification.

If the protocol is extended to S4-O.A to elicit more expressive epistemic states, however, 97.3 percent of responses will include Type-2 values. A Type-1 multi-criteria decision-making system facing this scenario must either clip the values to $[0, 1]$, introducing systematic bias and losing information, or reject the responses as protocol errors, discarding the richest epistemic expressions. Type-k Neutrosophic Sets provide the formal framework to avoid both failure modes: the extended responses are mathematically legitimate Type-2 elements, and the recursive SVNWA of Definition 3.6 aggregates them correctly while remaining backward compatible with Type-1 results by Proposition 3.1.

The practical recommendation for industrial multi-criteria decision-making practitioners is therefore threefold. First, for low-stakes decisions where Type-1 precision is sufficient, use constrained protocols (S1 or S4) and the standard SVNWA. Second, for high-stakes decisions requiring maximal epistemic expressiveness, use extended protocols (S4-O.A or S4-O.C) and the Type-2 SVNWA. Third, for audit and certification purposes, report both the Type-1 and Type-2 aggregated scores in order to quantify the information lost in the Type-1 projection. Chapter 11 develops this unified auditing protocol in detail.

3.7 Connections to Existing Frameworks

The Type-k framework relates to several existing extensions of the classical neutrosophic apparatus. Refined neutrosophic logic, introduced by Smarandache in 2013, increases cardinality at a single level by splitting T into sub-components (T_1, ..., T_p) that remain scalars. The Type-k extension increases recursion depth: at each level, each component is itself a full triplet. These two extensions are orthogonal and can be combined. A Refined Type-2 Neutrosophic Set would split each of the nine Type-2 scalars into sub-components, enabling simultaneous modelling of multiple sources of indeterminacy at multiple recursion depths. Plithogenic sets, introduced by Smarandache in 2018, assign Type-1 triplets per attribute; this is a structure that is Type-2 along the attribute dimension and is developed in detail in Chapter 5. SVN-based tensor extensions, developed in Chapter 4, provide the multi-mode array structure for scaling to large multi-criteria problems.

The Absorption Problem, identified by Mason in 2026, is the empirical observation that several language models map distinct epistemic states (paradox, ignorance, contingency) to the same scalar (T = 0, I close to 1, F = 0), collapsing the information content of these distinct phenomena into a single maximum-uncertainty value. Within the Type-k framework, the Absorption Problem is formally characterised as a forced Type-2 to Type-1 projection: the sub-structure of I (specifically, whether high indeterminacy reflects genuine ignorance with low F_I or structural conflict with high F_I) is lost when I is projected to its scalar Type-1 value. Type-2 representation preserves this distinction, enabling downstream reasoning systems to respond differently to the two causes of high indeterminacy. The plithogenic embedding developed in Chapter 5 provides an alternative resolution of the same problem through the attribute dimension.

3.7.3 N-alethic Neutrosophic Logic: The Perspectival Extension

Type-k Neutrosophic Sets extend the classical framework along what we call the vertical dimension: increasing recursion depth so that each component (T, I, F) is itself characterised by a full neutrosophic triplet at the next level. A complementary horizontal extension is provided by N-alethic Neutrosophic Logic , which indexes valuations by a perspective π and a time t rather than deepening the recursion of a single valuation. The formal definition introduces a perspectival valuation $V(\pi, t, A) = v_\pi^t(A) = (T_\pi^t(A), I_\pi^t(A), F_\pi^t(A))$ in $[0,1]^3$, where π ranges over a preordered set $(\Pi, \leq)$ of interpretive positions and t over a discrete time index. A type-k n-

alethic valuation combines both dimensions: $V_{\pi}^{t}(A) \in [0,1]^{3^k}$, giving an object that is simultaneously recursive (depth k) and perspectival (indexed by π and t).

The connection between the two frameworks runs deeper than a simple product. Consider the Type-2 element for a formula A, with its nine scalars organised as three sub-triplets $((T_T, I_T, F_T)$, (T_I, I_I, F_I), $(T_F, I_F, F_F))$. Now consider an n-alethic model with exactly three perspectives named pi_T, pi_I, and pi_F, each holding a single-valued neutrosophic valuation of A. The resulting 3×3 array $((T_{\pi_T}, I_{\pi_T}, F_{\pi_T})$, $(T_{\pi_I}, I_{\pi_I}, F_{\pi_I})$, $(T_{\pi_F}, I_{\pi_F}, F_{\pi_F}))$ is formally isomorphic to the Type-2 element. Under this isomorphism, the sub-triplet for the T-component in Type-2 corresponds to the valuation from the truth-perspective pi_T: the perspective that attends primarily to confirmatory evidence. The sub-triplet for the I-component corresponds to the indeterminacy-perspective pi_I: the perspective comfortable with sustained epistemic suspension. Type-k is thus a special case of n-alethic logic in which the perspectives are defined not by external agents but by the recursive structure of the valuation itself.

This isomorphism has a direct empirical reading. The per-vendor Type-2 rates documented in Table 3.2 (19.6 percent for Alibaba to 29.9 percent for Mistral) are, under the n-alethic interpretation, measurements of inter-perspectival conflict: each vendor constitutes a distinct perspective pi_i with its own force vector phi_{pi_i} in n-alethic terminology, and the Type-2 requirement rate is a proxy for the magnitude of that force. The gap between the S1 protocol (0 percent Type-2) and the S4-O.A protocol (97.3 percent) reflects the protocol's effect on the visibility of inter-perspectival conflict: constrained protocols suppress the conflict by forcing all components to $[0, 1]$; extended protocols allow it to surface. The n-alethic framework provides the formal vocabulary to articulate this contrast: S1 imposes $phi_{\pi} = 0$ for all perspectives, while S4-O.A releases the force constraint and the latent perspectival disagreement becomes empirically visible.

The research agenda that this isomorphism opens is discussed in Chapter 12 (Problem 11). The formal development of N-alethic Neutrosophic Logic, including the perspective space, the transition operator, the non-synthetic stability axiom, and the perspectival consequence relation, is the subject of a companion paper currently in preparation.

3.8 Conclusion

This chapter introduced Type-k Neutrosophic Sets as a formal recursive extension of classical neutrosophic logic in which each component is itself characterised by a full neutrosophic triplet. The strict expressive hierarchy of Theorem 3.1 establishes that each Type-k level is strictly more expressive than its predecessor, and the backward-compatibility result of Proposition 3.1 ensures that existing Type-1 multi-criteria decision-making models can be extended to Type-k without redesign. The empirical results across 2,419 evaluations of six foundation models provide the first quantitative evidence that Type-2 representation is operationally necessary when extended elicitation protocols are used: 24.7 percent of all responses and 97.3 percent under the overset protocol require Type-2 representation, and the cross-vendor consistency (19.6 to 29.9 percent) rules out model-specific artefacts and confirms a structural property of the evaluation landscape.

The recursive Type-k construction provides a vertical dimension of refinement: nesting the neutrosophic triplet within itself. The next part of the book develops a horizontal dimension: tensorial extensions that distribute the triadic structure across the modes of a multi-criteria evaluation. Chapters 4 through 6 introduce single-valued neutrosophic tensors, plithogenic tensors, and the decision-theoretic consequences of the resulting five-level hierarchy.

PART II

Tensorial Extensions

Chapter 4

SVN Tensors: From Vectors to Multilinear Structures

This chapter introduces single-valued neutrosophic tensors (SVN tensors) as the multilinear-algebraic generalisation of the classical fuzzy tensors of Chen and Lu (2019) and the intuitionistic fuzzy tensors of Chen and Chen (2019). The two prior frameworks are inadequate for representing the joint structure of high-dimensional, multi-criteria evaluation data when the residual hesitation cannot be reduced to a single derived quantity. SVN tensors lift the entries of a tensor from scalar membership degrees to independent triplets (T, I, F), enabling Tucker decomposition, cut-tensor representation, and Einstein contraction to operate at the level of full neutrosophic objects rather than projected scalars. We prove a strict hierarchy theorem connecting fuzzy, intuitionistic-fuzzy, and SVN tensors, establish that SVN contraction is a conservative extension of the recently introduced Tensor Logic of Domingos (2025), and present an empirical application to the ethical evaluation of large language models that comprises 672 evaluations across four models, twelve dilemmas, and four protocols of increasing structural richness.

4.1 Motivation: Why Tensors

A Type-k neutrosophic evaluation matrix of m alternatives by n criteria has the natural mathematical structure of a tensor of order at least three, with the additional axes encoding either the components of the neutrosophic triplet, the recursion depth, the evaluator identity, or all three simultaneously. When the data are stored as a flat table of scalars, the multilinear structure that connects the axes is lost; when they are stored as a tensor, decomposition methods such as Tucker and Higher-Order Singular Value Decomposition can extract low-rank approximations that respect the structure. The computational economy of tensor methods is well established in machine learning, where they underlie recommender systems, knowledge graph completion, neural network weight compression, and the recent Tensor Logic programme of Domingos for unifying symbolic and neural computation.

The challenge is that the entries of these tensors are typically real scalars in $[0, 1]$ or in the real line. The empirical chapters of this book demonstrate that the epistemic states reported by modern language models cannot be faithfully represented by scalars: they require independent

triplets, and frequently require triplets whose components exceed the unit interval. Lifting the entries of a tensor to single-valued neutrosophic triplets is therefore the natural generalisation of the tensor methods of classical machine learning to the epistemic regime that this book describes.

4.2 Definition of SVN Tensors

We adopt the tensor notation of Kolda and Bader (2009). Tensors are denoted by calligraphic uppercase letters such as T; the entry of a tensor T at position (i_1, ..., i_n) is T(i_1, ..., i_n); the mode-k product with a matrix U is denoted T x_k U; the Hadamard (elementwise) product is denoted by the operator circle-dot; and Einstein summation over a repeated index is implicit in tensor contraction.

Definition 4.1 (Single-valued neutrosophic tensor). A single-valued neutrosophic tensor of order n on dimensions $\mathrm{I}_1 \times \ldots \times \mathrm{I}_\mathrm{n}$ is a triple , $\mathcal{T} = (\mathcal{T}_\mathcal{T}, \mathcal{T}_\mathcal{I}, \mathcal{T}_\mathcal{F})$, $\mathcal{T}_\mathcal{X}: [\mathrm{I}_1] \times \cdots \times [\mathrm{I}_\mathrm{n}] \to [0,1]$, $\mathrm{X} \in \{\mathrm{T}, \mathrm{I}, \mathrm{F}\}$. where $\mathcal{T}_\mathcal{T}, \mathcal{T}_\mathcal{I}, \mathcal{T}_\mathcal{F}$ denote the truth-membership, indeterminacy-membership, and falsity-membership tensor components, respectively. For every index tuple $\mathrm{i} = (\mathrm{i}_1, \ldots, \mathrm{i}_\mathrm{n}) \in [\mathrm{I}_1] \times \cdots \times [\mathrm{I}_\mathrm{n}]$, the SVN admissibility condition is given by $0 \leq \mathcal{T}_\mathcal{T}(\mathrm{i}) + \mathcal{T}_\mathcal{I}(\mathrm{i}) + \mathcal{T}_\mathcal{F}(\mathrm{i}) \leq 3$.No normalization condition such as $\mathcal{T}_\mathcal{T}(\mathrm{i}) + \mathcal{T}_\mathcal{I}(\mathrm{i}) + \mathcal{T}_\mathcal{F}(\mathrm{i}) = 1$ is imposed.

Definition 4.2 (SVN tensor operations). *The SVN sum, scalar product, Hadamard product, and mode-k matrix product on SVN tensors are defined componentwise, using the t-conorm-product and t-norm-product operators of Ye (2014). The SVN Einstein contraction over a repeated mode contracts each component using the Hadamard SVN product followed by the SVN sum.*

4.3 Tucker Decomposition for SVN Tensors

The Tucker decomposition is the cornerstone of multilinear algebra for tensors of order three and higher. It generalises the singular value decomposition of matrices by representing a tensor as the mode-wise product of a core tensor with factor matrices, one per mode. For SVN tensors, the decomposition is performed independently on each of the three components.

Theorem 4.1 (Tucker decomposition for SVN tensors). *Every SVN tensor T admits a Tucker decomposition componentwise: there exist a core SVN tensor G and factor matrices*

$U_X^{(k)}$ for k = 1, ..., n and X in {T, I, F} such that T_X is approximately $G_X\ x_1\ U_X^{(1)}\ x_2 \ldots x_n\ U_X^{(n)}$ for each component X. The reconstruction satisfies the SVN constraint pointwise.

Proof. Apply the standard Tucker decomposition (Kolda and Bader 2009, Theorem 4.2) to each real component $T_X \in \mathbb{R}^{I_1 \times \cdots \times I_n}$, with truncation projected back into $[0,1]$. Pointwise the constraint $T + I + F \leq 3$ is preserved trivially since each component lies in $[0,1]$. ■

Theorem 4.2 (Cut-tensor representation). *Let α, β, γ in $[0,1]$ with $\alpha + \beta + \gamma$ at most 3. Define cut sets $[T]_\alpha^T = \{i: T_T(i) \geq \alpha\}$, $[T]_\beta^I = \{i: T_I(i) \geq \beta\}$, $[T]_\gamma^F = \{i: T_F(i) \leq \gamma\}$. Then T is recoverable from its cut family.*

Proof. Each component independently satisfies the fuzzy-cut representation theorem of Chen and Lu (2019, Theorem 2). The independence of T, I, and F in the SVN tensor setting permits componentwise treatment, and the recovery from cuts is obtained by intersection of the level sets in the standard way. ■

4.4 The Hierarchy Theorem

The most important structural result of this chapter is the strict hierarchy that connects fuzzy tensors, intuitionistic fuzzy tensors, and SVN tensors. This hierarchy parallels the strict hierarchy of Type-k Neutrosophic Sets established in Chapter 3 and, when extended in Chapter 5 to include plithogenic tensors, produces the complete four-level expressivity hierarchy that underlies the decision-theoretic results of Chapter 6.

Theorem 4.3 (Strict hierarchy). *There exist canonical inclusions FuzzyTensor into IFTensor into SVNT, and each inclusion is strict.*

Proof. The inclusions are explicit. A fuzzy tensor T_F embeds as the IFT (T_F, 1 - T_F). An IFT (μ, ν) with $\mu + \nu$ at most 1 embeds as the SVN tensor (μ, 1 - μ - ν, ν). For strictness, we exhibit witnesses. FuzzyTensor $\subsetneq$ IFTensor: the IFT (0.4, 0.4) has $\mu + \nu = 0.8 < 1$ with hesitation 0.2 that no fuzzy tensor encodes (Atanassov 1986). IFTensor $\subsetneq$ SVNT: the SVN tensor entry (0.05, 0.20, 0.92) has T + I + F = 1.17 > 1 and is therefore not expressible as any IFT under the constraint $\mu + \nu$ at most 1. We refer to the regime $T + I + F > 1$ as ethical hyper-truth; it occurred in 100 percent of high-stakes ethical cases evaluated by Claude Sonnet 4.6 and GPT-4o in the experiment of Section 4.7. ■

4.5 SVN Contraction as a Conservative Extension of Tensor Logic

Domingos (2025) recently introduced Tensor Logic as a unified language for neural and symbolic artificial intelligence, reducing logical inference to Einstein contraction over scalar tensors in $[0, 1]$ or $\{0, 1\}$. Tensor Logic is silent on independent indeterminacy and on attribute interdependence; it operates entirely in the fuzzy regime where falsity is the complement of truth and indeterminacy is absent. The next theorem establishes that SVN tensor contraction is the natural lift of Tensor Logic to the regime in which evidence is partial or conflicting.

> **Theorem 4.4 (Conservative extension of Tensor Logic).** *Let $\mathcal{T}$ be an SVN tensor with T_I identically zero and T_F identically equal to 1 - T_T pointwise. Then for any compatible SVN tensor S in the same regime, the SVN Einstein contraction $\mathcal{T}$ contracted with S reduces exactly to the classical Tensor Logic contraction of Domingos (2025) on the truth components.*
>
> *Proof.* In the regime I identically zero, the indeterminacy component of every operation evaluates to zero by the definition of the t-norm-product on the indeterminacy axis. In the regime F identically equal to 1 - T, the falsity component is determined by the truth component, so the SVN tensor collapses to a single degree of freedom per cell, isomorphic to the fuzzy tensor in $[0, 1]$. The SVN sum reduces to the t-conorm-product, which is precisely the operator that Domingos uses for the soft-OR semantics of Datalog-as-tensor. The composite contraction therefore coincides cell-by-cell with the Tensor Logic contraction. ■

Theorem 4.4 is the anchoring result of this chapter: SVN tensors and the plithogenic tensors of Chapter 5 do not compete with Tensor Logic; they extend it conservatively to a regime in which evidence is partial or conflicting, which is precisely the regime in which large language models operate when asked to evaluate ethical dilemmas or contested factual claims. The conservative extension property has practical consequences for downstream pipelines: a system that processes scalar tensors via Tensor Logic can be lifted to process SVN tensors without modification of the contraction logic; only the data representation changes.

4.6 Robustness under Perturbation

An important property of any aggregation operator that incorporates expert-elicited inputs is its robustness to small perturbations in those inputs. For SVN tensors, the relevant perturbation is in the entries themselves; for the plithogenic tensors of Chapter 5, it will be in the contradiction

function. We state the robustness result here for SVN tensors with respect to elementwise perturbation, deferring the corresponding result for plithogenic contraction to Chapter 5.

Theorem 4.5 (Stability of SVN contraction). *Let $\mathcal{T}$ and $\tilde{\mathcal{T}}$ be two SVN tensors with the same dimensions, and suppose that $\left\|\mathcal{T} - (\widetilde{\mathcal{T}})\right\|_\infty \leq$ componentwise on each of the T, I, F components. Then for any compatible SVN tensor S, the contraction (denoted $*$) satisfies $\left\|\mathcal{T} * S - (\widetilde{\mathcal{T}}) * S\right\|_\infty \leq \delta\|S\|_\infty$, with the bound holding componentwise. In compact form:* $\left\|\mathcal{T} - (\widetilde{\mathcal{T}})\right\|_\infty \leq \delta \Rightarrow \left\|\mathcal{T} * S - (\widetilde{\mathcal{T}}) * S\right\|_\infty \leq \delta\|S\|_\infty$

Proof. Each component of the SVN contraction is a t-conorm-product or t-norm-product aggregation of Hadamard products. Both t-conorm-product and t-norm-product are Lipschitz with constant at most 1 on the unit cube, and the Hadamard product with a tensor of pointwise bound $\|S\|_\infty$ scales the perturbation by at most $\|S\|_\infty$. The composition therefore propagates the input perturbation with the stated bound. ■

4.7 Application: Ethical Evaluation of Large Language Models

We apply the SVN tensor framework to the ethical evaluation of four large language models: GPT-4o, Claude Sonnet 4.6, Gemini 2.5 Pro, and Llama 3.1 70B Instruct, accessed through a unified OpenRouter endpoint. Twelve ethical dilemmas spanning six categories were evaluated under four protocols of increasing structural richness, with three repetitions per (case, model, protocol) cell, yielding 576 evaluations. A separate paired contrastive experiment of 96 evaluations probed model behaviour under Global-North versus Global-South case formulations. Sampling temperature was 0.7; maximum tokens were 4096 to 8192. The total dataset comprises 672 evaluations.

4.7.1 Protocols and Attributes

The four protocols correspond to four levels of representational structure. P1 (binary) returns one of three categorical labels: PERMIT, NO_PERMIT, ESCALATE. P2 (scalar) returns a real value in $[0, 1]$ interpreted as a fuzzy ethical score. P3 (neutrosophic global) returns a single triplet (T, I, F) in $[0,1]^3$ representing the global epistemic state of the evaluator. P4 (plithogenic tensor)

returns a full per-attribute triplet for each of fourteen ethical attributes plus a contradiction matrix over the attribute pairs.

The fourteen ethical attributes follow established AI ethics frameworks (Floridi et al. 2018; Jobin et al. 2019; Hendrycks et al. 2021): Safety, Non-maleficence, Fairness, Privacy, Transparency, Explainability, Human autonomy, Accountability, Social benefit, Human dignity, Cultural inclusion, Sustainability, Proportionality, and Traceability. Each attribute is evaluated independently by the language model, and the contradiction matrix is elicited in the same response.

4.7.2 The Plithogenic Tensor of an Evaluation

Each P4 evaluation produces a tensor X[s, m, g, a, c] indexed by case s, evaluating model m, affected group g (implicit per case), ethical attribute a, and neutrosophic component c in {T, I, F}. Two distinct structural objects coexist in each evaluation. The SVN tensor body T = (T_T, T_I, T_F) is a 14 by 3 array per (s, m, g) triple, for a total of 12 times 4 times 14 times 3 = 2,016 cells across the experiment. The contradiction matrix C[s, a_i, a_j] is one symmetric 14 by 14 matrix per case (with zero diagonal), yielding 91 unique off-diagonal entries per case, or 12 times 91 = 1,092 unique contradiction values across the experiment. Together, the two objects encode 3,108 numbers per protocol. The scalar P2 protocol, by contrast, produces 12 times 4 = 48 numbers: exactly one ethical score per (case, model) pair, with no internal structure. The plithogenic protocol preserves 64 times more structural information per case than the scalar protocol.

4.7.3 Three Findings Inaccessible to Scalar Protocols

Three structural findings emerge from the P4 evaluations that no scalar protocol could recover. We summarise them here and develop the decision-theoretic implications in Chapter 6.

Finding 1: three alignment regimes, not two. Under the P4 protocol, the recommendation distribution reveals three structurally distinct regimes. Llama 3.1 70B and Gemini 2.5 Pro are bipolar: every case falls into a single output bucket (HUMAN_REVIEW for Llama at 100 percent of 36 cases, BLOCK for Gemini at 100 percent). GPT-4o and Claude Sonnet 4.6 span multiple buckets, and Claude is the only model that uses three of the four output levels. Under the scalar P2 protocol, the four models are statistically indistinguishable on these cases (mean ethical scores range from 0.14 to 0.31, all low). Only the P4 protocol surfaces the three-regime structure.

Table 4.1 reports the recommendation distribution per model.

Model	PERMIT	WARN	HUMAN_REVIEW	BLOCK
Llama 3.1 70B	0	0	36 (100%)	0
GPT-4o	1	0	30 (83%)	5
Claude Sonnet 4.6	0	1	13	22 (61%)
Gemini 2.5 Pro	0	0	0	35 (100%)

Table 4.1. Recommendation distribution under P4 across four LLMs (n = 36 cases per model).

Finding 2: human autonomy is the structural central node of the contradiction graph. Of the seven most frequent inter-attribute contradictions surfaced across all 143 P4 evaluations, four involve human autonomy. The empirical centrality of this attribute is, to our knowledge, the first quantitative measurement of a thesis advanced informally in the AI alignment literature: that the central problem of governing artificial intelligence is not any single value but the structural tension between user autonomy and every gatekeeping principle designed to protect users from themselves or from others. The dominant contradictions are Human autonomy versus Safety (frequency 86, mean contradiction degree 0.73), Privacy versus Transparency (frequency 85, mean degree 0.70), Human autonomy versus Non-maleficence (frequency 23, mean degree 0.79), Fairness versus Social benefit (frequency 23, mean degree 0.77), Accountability versus Human autonomy (frequency 22, mean degree 0.65), Non-maleficence versus Social benefit (frequency 20, mean degree 0.75), and Privacy versus Safety (frequency 19, mean degree 0.83).

Finding 3: plithogenic metrics preserve more variance than scalar metrics. The variance of model outputs across the same set of cases differs substantially between protocols. The P2 scalar protocol yields a mean of 0.251 with standard deviation 0.243 and range [0.00, 0.90]. The P4 plithogenic metrics (Epistemic Information Load, Ethical Risk Level, Ethical Contradiction Index) yield comparable variance but are informationally orthogonal to each other and to the recommendation tier. The P2 protocol collapses three independent structural quantities (indeterminacy, risk, and inter-attribute conflict) into a single number, a flattening that the plithogenic protocol explicitly preserves.

4.7.4 Per-Model Contradiction-Detection Capacity

Independently of the recommendation regime, the four models differ strongly in the richness of contradictions they surface. Claude Sonnet 4.6 surfaces an average of 6.6 contradictions per case with mean Ethical Contradiction Index 0.728. Gemini 2.5 Pro surfaces 4.5 per case with mean ECI 0.797. GPT-4o surfaces 2.9 per case with mean ECI 0.711. Llama 3.1 70B surfaces 2.1 per case with mean ECI 0.630. Claude detects 3.1 times as many contradictions per case as Llama, despite both models producing tensors of identical dimensionality. Detecting that two attributes are in tension is therefore a capability distinct from scoring each attribute, and it is recovered only under the plithogenic protocol.

4.8 Conclusion

SVN tensors lift the entries of a classical multilinear object from scalar membership degrees to independent neutrosophic triplets. They admit Tucker decomposition, cut-tensor representation, and Einstein contraction; they conservatively extend the Tensor Logic of Domingos (2025) to the regime of partial and conflicting evidence; and they yield a strict expressivity hierarchy FuzzyTensor strictly contained in IFTensor strictly contained in SVNT. The empirical application to 672 evaluations of large language model ethical reasoning confirms that the regime $T + I + F > 1$, inaccessible to intuitionistic-fuzzy representations, is realised in 100 percent of high-stakes ethical cases by the most capable models in the panel. Chapter 5 extends the framework one further level by introducing an explicit contradiction function over attribute pairs, yielding plithogenic tensors.

Chapter 5

Plithogenic Tensors: A Hierarchical Multilinear Framework

Plithogenic tensors enrich SVN tensors with an explicit contradiction function over pairs of attributes. This chapter develops the formal apparatus, proves the embedding theorem that situates Mason's S4 declared-loss protocol as a faithful instance of single-valued plithogenic neutrosophic evaluation, reformulates the Absorption Problem as a non-injective scalar projection from a richer object, and presents the multi-vendor empirical study (n = 1,830 main evaluations plus 90 controls plus 250 cross-evaluator cells, totalling 2,170 evaluations) that grounds the framework in current foundation-model behaviour. The chapter closes with a detailed comparison against six alternative frameworks (Dempster-Shafer evidence theory, type-2 fuzzy sets, modal logic, rough sets, case-based reasoning, and Atanassov intuitionistic fuzzy logic) and an account of the auditor-effect finding that gpt-4o is structurally isolated in projecting offset evaluations back onto the unit cube.

5.1 Plithogenic Structures: Background

Plithogeny, introduced by Smarandache in 2018, extends fuzzy, intuitionistic-fuzzy, and neutrosophic sets by attaching to each attribute a contradiction degree relative to a designated dominant attribute. The formal apparatus is the five-tuple plithogenic structure.

> **Definition 5.1 (Plithogenic structure).** *A plithogenic structure is a five-tuple P = (P, v, V, d, c), where P is a set of plithogenic elements (in the language-model evaluation setting, statements under evaluation); v is the dominant attribute; V is the spectrum of attribute values (finite or continuous); d : P x V to* $[0,1]^n$ *is the degree-of-appurtenance function (with n = 1 fuzzy, n = 2 intuitionistic, n = 3 neutrosophic); and c : V x V to* $[0,1]$ *is the contradiction function, satisfying c(v, v) = 0 and c(*v_i*,* v_j*) = c(*v_j*,* v_i*).*

> **Definition 5.2 (Single-valued plithogenic neutrosophic set).** *A single-valued plithogenic neutrosophic set is a plithogenic structure with n = 3 and d(p, v) = (*$T_v(p)$*,* $I_v(p)$*,* $F_v(p)$*) in* $[0,1]^3$ *for each p in P and v in V.*

The contradiction function c is the formal device that distinguishes plithogeny from earlier multi-valued logics. Two attributes with c(v_i, v_j) = 0 are taken as semantically synonymous; with c = 1 they are taken as maximally opposed. Operationally, c may be estimated by lexical overlap

(Jaccard or token cosine), by sentence-embedding distance under a contemporary encoder, or by domain-expert elicitation. The choice of estimator is itself a methodological decision that should be documented in any application.

5.2 Plithogenic Tensors: Definition and Operations

Definition 5.3 (Plithogenic tensor). *A plithogenic tensor of order n is a tuple T_P = (T, A, a_D, c), where T is an SVN tensor of order n in which one mode (say mode k-star) is indexed by an attribute set A = {a_1, ..., a_m}; a_D in A is the dominant attribute; and c : A x A to* $[0, 1]$ *is a symmetric contradiction function with c(a_D, a_D) = 0.*

Definition 5.4 (Plithogenic Einstein contraction). *For plithogenic tensors T_P and S_P contracted over a repeated attribute mode, the plithogenic contraction is (T_P contracted with S_P) = $oplus_j$ (1 - c(a_j, a_D)) times (T_X Hadamard S_X)_..., j, ..., performed componentwise for X in {T, I, F}. The weight 1 - c(a_j, a_D) attenuates the contribution of attributes that are maximally opposed to the dominant attribute.*

5.3 Embedding of Mason's S4 into the Plithogenic Framework

Mason (2026) introduced the S4 protocol, in which an evaluator returns not only a global (T, I, F) triplet but also a list of declared losses, each of the form (what, why, severity) with severity in $[0, 1]$. Mason's empirical claim is that the declared losses recover epistemic distinctions that the scalar (T, I, F) cannot express, in particular the distinction between paradox, ignorance, and contingency when the scalar projection collapses to (T = 0, I close to 1, F = 0). We provide the formal foundation for this claim by exhibiting a faithful injective embedding of S4 into the single-valued plithogenic neutrosophic structure.

Theorem 5.1 (Faithful plithogenic embedding of Mason's S4). *Let S^{S4} denote the set of all S4 outputs. Define φ : S^{S4} to P^{SVPN} by φ(M(s)) = (P_s, v, V_s, d_s, c_s) with P_s = {s}, v = epistemic_limitation, V_s = {$what_1$, ..., $what_k$}, d_s(s, $what_j$) = $(T, I \times \text{severity}_j, F)$, and c_s($what_i$, $what_j$) = 1 - J(why_i, why_j), where J denotes the Jaccard token-overlap similarity. Then (i) φ is well-defined; (ii) φ is injective; (iii) the scalar projection π satisfies π(φ(M)) = (T, I, F); (iv) the plithogenic conjunction and disjunction are closed on the image φ(S^{S4}).*

Proof. Well-definedness is immediate from the definitions: each S4 output produces a unique tuple with the stated structure, and the contradiction function is well-defined since Jaccard is symmetric, non-negative, bounded by 1, and zero on identical sets. Injectivity follows from the regularity condition (Mason 2026, Section 4) that at least one declared loss has severity 1 (the saturating loss); two distinct S4 outputs differ either in (T, I, F) or in the list of losses, and the embedding preserves both pieces of information. The scalar projection follows from summing the per-attribute triplets weighted by 1 - c($what_j$, $what_D$) and recovering (T, I, F) under the convention that the dominant what corresponds to the saturating loss. Closure under plithogenic conjunction and disjunction is verified by direct calculation: the operators preserve the attribute set V_s and respect the contradiction function. The full four-part proof is in the accompanying technical note. ■

Corollary 5.1 (Plithogenic resolution of the Absorption Problem). *Two S4 outputs whose scalar projections coincide but whose loss vocabularies are disjoint map to non-isomorphic plithogenic structures, and the contradiction function assigns maximal distance 1 between cross-vocabulary attributes.*

Corollary 5.2 (S4-N as natural plithogenic extension). *The image $\varphi(S^{S4})$ is a strict subset of the space of single-valued plithogenic neutrosophic sets. The relative complement contains every structure with non-trivially varying per-attribute T_v and F_v. We refer to the protocol that elicits these richer outputs as S4-N.*

5.4 The Absorption Problem Revisited

The Absorption Problem in Mason's terminology is the empirical observation that several language models map distinct epistemic states (paradox, ignorance, contingency) to the same scalar (T, I, F) close to (0, 1, 0). The scalar projection collapses the distinction. We make two distinct claims about this phenomenon. The structural claim is mathematical: the projection π is non-injective, so two distinct plithogenic structures may collapse to the same scalar. The empirical claim is that per-attribute decomposition under S4-N attenuates the absorption.

Proposition 5.1 (Structural non-collapse of plithogeny). *Let π : P to (T, I, F)_global be the scalar projection. Suppose p_1, p_2 in P satisfy $\pi(P_{p_1}) = \pi(P_{p_2})$ but V_{p_1} intersect V_{p_2} =*

empty. Then P_{p_1} and P_{p_2} are non-isomorphic plithogenic structures, and the unified plithogenic distance d_P of Definition 5.5 below is strictly positive.

Definition 5.5 (Unified plithogenic distance). *For two plithogenic structures P_1, P_2 over a common universe of statements,* $d_P(P_1, P_2) = \alpha d_{\text{scalar}} + \beta\big(1 - J(V_1, V_2)\big) + \gamma\overline{(c)}(V_1 \cup V_2)$ *where d_{scalar} is the Manhattan distance on the projected triples, J(V_1, V_2) is the Jaccard similarity of attribute spectra, and c-bar is the average contradiction across cross-set pairs. The weights satisfy α, β, γ at least 0 and $\alpha + \beta + \gamma = 1$.*

The unified plithogenic distance subsumes scalar Manhattan distance as the $\alpha = 1$ special case and is strictly more discriminating whenever $\alpha < 1$ and the attribute spectra differ. Empirically, Mason's reported scalar Manhattan distances of 0.034 (Claude), 0.040 (Llama), and 0.000 (Mistral) for paradox versus ignorance yield d_scalar near zero. The Jaccard similarities of 0.097 (Claude), 0.056 (Llama), and 0.066 (Mistral) yield attribute-distance contributions in [0.90, 0.94]. With balanced weights $\alpha = \beta = \gamma = 1/3$, the unified distance produces d_P at least 0.30 for every Absorption case in his data, far above any reasonable threshold for identical evaluation. The Absorption Problem is therefore an artefact of dropping the attribute and contradiction terms from the metric.

5.5 Empirical Validation: The Multi-Vendor Study

5.5.1 Design

Six vendors were tested via OpenRouter: Alibaba Qwen-3-235B, Anthropic Claude Sonnet 4, DeepSeek Chat, Meta Llama-4-Maverick, Mistral Medium-3.1, and OpenAI GPT-4o. Five protocols were elicited: S1 (classical neutrosophic on $[0,1]^3$), S4 (Mason 2026), S4-N (per-attribute neutrosophic), S4-O.A (overset on $[0,2]^3$), and S4-O.C (peer-evaluation offset on $[-1,2]^3$). Five epistemic phenomena were evaluated: Paradox, Ignorance, Vagueness, Ethical Contradiction, and Contingency. Ten repetitions were performed per cell at temperature 0.7. Three control conditions were included: prompt ablation (S1.v2 and S1.v3 on Paradox and Ethical), tautology control (S1 on three canonical tautologies), and uncertainty-quantification baselines (Kuhn semantic entropy and a strict reformulation of SelfCheckGPT). A crossover extension was added in which each non-self source vendor was audited under S4-O.C by three

peer evaluators, totalling 1,830 main evaluations plus 250 cross-evaluator cells plus 90 baseline evaluations after deduplication.

5.5.2 Hyper-Truth Replicates across All Six Vendors

Under the S1 protocol, the rate of hyper-truth ($T + I + F > 1$) ranges from 0.62 (Mistral) to 1.00 (Alibaba, Anthropic). The phenomenon-by-hyper-truth chi-square test is significant in every protocol ($p < 10^{-22}$). The phenomenon is structural, not OpenAI-specific.

5.5.3 The Tautology Control: A Clean Refutation of the Format-Artefact Hypothesis

Across the three canonical tautologies (2 + 2 = 4; all bachelors are unmarried; it is raining or it is not raining) and all six vendors at five repetitions each (90 cells total), the hyper-truth rate is exactly 0.000. This refutes the objection that hyper-truth is produced by the prompt structure rather than by the underlying epistemic phenomenon. When the underlying statement is determinate, the models report (T = 1, I = 0, F = 0); when the statement is paradoxical, vague, or contingent, they report hyper-truth. The protocol distinguishes the two regimes cleanly.

5.5.4 Plithogenic Decomposition (S4-N) Attenuates the Absorption Problem

Under S4-N, the indeterminacy mean drops on the two phenomena where Absorption is most acute: Ignorance from I = 0.953 (S4) to 0.825 (S4-N), and Paradox from I = 0.990 to 0.852. Truth and falsity coordinates rise (for Ignorance, T rises from 0.098 to 0.194; for Paradox, F rises from 0.108 to 0.443) as the previously collapsed epistemic mass redistributes across attributes. A McNemar test comparing S1 with S4-N indicates that 106 cells break the absorption while 37 reconverge ($p < 0.001$). However, 58.2 percent of cells under S4-N still produce $T + I + F > 1$: the redistribution does not collapse absorption, it transforms it into a structured distribution that the scalar projection can no longer flatten. The corresponding theoretical thesis is therefore reformulated from resolves the Absorption Problem to attenuates the Absorption Problem and decomposes it into interpretable plithogenic components.

5.5.5 The Auditor-Effect Crossover: The Central Empirical Finding

The peer-evaluation offset protocol S4-O.C admits a key experimental decomposition not available in the self-evaluation protocols: the auditor identity can be varied while the audited model is held fixed. We ran a strict 4 x 3 evaluator crossover: the four source vendors (Alibaba, Anthropic, DeepSeek, Mistral) were audited by all three peer evaluators (gpt-4o, claude-sonnet-4,

llama-4-maverick) at 200 cells per evaluator. The rate of extended-range usage, defined as (T < 0) or (F < 0) or (T > 1) or (I > 1) or (F > 1), differs sharply across auditors.

Auditor	**Extended-range usage**	**% T < 0**	**% F < 0**
claude-sonnet-4	0.490	7.0%	7.0%
llama-4-maverick	0.515	7.0%	6.0%
gpt-4o	0.120	0.0%	0.0%

Table 5.1. Auditor-effect crossover on S4-O.C. Rate of extended-range usage by evaluator.

Pearson chi-square three-way test yields chi-square = 83.50, df = 2, p = 7.4 x 10^{-19}. Pairwise comparisons show that Claude versus Llama is statistically indistinguishable (chi-square = 0.16, p = 0.69); both differ massively from gpt-4o (Claude versus gpt-4o p = 2.2 x 10^{-15}; Llama versus gpt-4o p = 5.4 x 10^{-17}). gpt-4o is structurally isolated in projecting the S4-O.C audit back onto $[0, 1]$; Claude and Llama-4, despite their architectural differences, exercise the offset paradigm to a statistically indistinguishable extent. Negative coordinates are produced by Claude in approximately 7 percent of cells, by Llama-4 in approximately 6 to 7 percent, and by gpt-4o in 0 percent: a clear gradient that documents a structural property of the auditor rather than the audited model.

5.5.6 Plithogenic Uncertainty is Orthogonal to Probability-Based UQ

We computed Pearson correlations between the S1 hyper-truth rate (aggregated by vendor by phenomenon, n = 30) and two mainstream uncertainty-quantification baselines: Kuhn-Gal-Farquhar semantic entropy (2023) and a strict reformulation of SelfCheckGPT (Manakul et al. 2023). Both correlations are non-significant: r = -0.10 (p = 0.60) for hyper-truth versus semantic entropy and r = -0.12 (p = 0.53) for hyper-truth versus SelfCheckGPT-strict. The magnitudes are small, supporting the claim that plithogenic uncertainty captures a dimension orthogonal to probability-bounded uncertainty quantification. The two paradigms measure complementary aspects: semantic entropy and SelfCheckGPT operationalise uncertainty as variability across stochastic generations of an answer, while the neutrosophic protocols operationalise uncertainty as a self-reported epistemic stance about a single statement.

5.6 Comparison against Alternative Frameworks

A natural objection to the plithogenic embedding is that other multi-valued or multi-attribute frameworks could equally accommodate the declared-loss structure. We address this objection by establishing six criteria, motivated by the S4 protocol itself, and showing that only plithogeny satisfies their conjunction. The criteria are: C1, hyper-truth admissibility ($T + I + F > 1$); C2, explicit attribute spectrum (first-class structured set V); C3, contradiction function on attribute values; C4, compositional closure under intersection, union, negation; C5, graded[0,1][3] membership rather than binary classification; and C6, connection to dynamic epistemic state (time-indexed extension).

Framework	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**
Plithogenic neutrosophic	yes	yes	yes	yes	yes	yes
Dempster-Shafer evidence	no	no	no	yes	yes	partial
Type-2 fuzzy sets	no	no	no	yes	yes	no
Modal logic (S4, S5)	n/a	no	no	yes	no	yes
Rough sets	no	partial	no	yes	no	no
Case-based reasoning	no	partial	no	no	no	partial
Atanassov intuitionistic	no	no	no	yes	yes	no

Table 5.2. Plithogenic neutrosophic logic versus six alternative frameworks across six criteria. Only plithogenic neutrosophic logic satisfies C1 through C6 simultaneously.

5.7 Beyond the Unit Cube: Overset, Underset, and Offset

Smarandache (2016) generalised neutrosophic structures to three regimes that extend the unit cube and are particularly relevant to the empirical audit of language model outputs. A neutrosophic evaluation lies in the overset regime if at least one component exceeds 1; in the underset regime if at least one component falls below 0; and in the offset regime if both occur simultaneously across the three components.

The motivating real-world examples in Smarandache's exposition are striking in their simplicity. An employee who works overtime deserves a degree of membership in the company strictly greater than 1 with respect to a regular full-time employee whose membership is normalised to 1; conversely, an employee whose net contribution is negative deserves a degree of membership strictly less than 0. The standard fuzzy and neutrosophic restriction to $[0, 1]$ cannot represent either situation. The empirical detection of overset and underset regimes in language model output is non-trivial because models trained on probabilistic targets tend to clamp outputs to $[0, 1]$ regardless of the prompt specification, but only some models. The auditor-effect finding of Section 5.5.5 is in this sense a discovery of which models clamp and which exercise the extended range: gpt-4o clamps, while claude-sonnet-4 and llama-4-maverick do not.

5.8 Conclusion

Plithogenic tensors extend SVN tensors with an explicit contradiction function on attribute pairs. Theorem 5.1 establishes that Mason's S4 declared-loss protocol is a faithful instance of single-valued plithogenic neutrosophic evaluation. The Absorption Problem dissolves at the structural level (Proposition 5.1) and attenuates at the empirical level under per-attribute decomposition. The multi-vendor empirical study confirms that hyper-truth replicates across six vendors, that the tautology control rules out the format-artefact hypothesis, that per-attribute decomposition attenuates absorption in a McNemar-significant way, and that the auditor identity (not the audited model identity) controls the use of the extended $[-1,2]$ range under the peer-evaluation offset protocol. Chapter 6 develops the decision-theoretic consequences of the five-level hierarchy in the context of a clinical decision support benchmark.

Chapter 6

Decision Optimality under Epistemic Indeterminacy

This chapter develops the decision-theoretic consequences of the tensor hierarchy established in Chapters 4 and 5. We establish the five-level hierarchy Crisp strictly contained in Fuzzy strictly contained in IFT strictly contained in SVN strictly contained in Plithogenic, prove the Decision Optimality theorem (Theorem 6.4) and the Contradiction Visibility theorem (Theorem 6.5), and apply the resulting framework to a clinical decision support benchmark comprising three language-model-based systems, four epistemic criteria, and three domain experts. The chapter closes with a comparison against AHP and TOPSIS, the two dominant multi-criteria decision-making methods of the operations-research literature, showing that the plithogenic tensor evaluation produces a substantially richer report along nine dimensions of output that scalar methods cannot recover.

6.1 The Five-Level Hierarchy

We summarise the entry structure, constraints, and canonical score function for each of the five tensor classes in Table 6.1.

Class	Entry	Constraint	Params/entry	Score S(.)
Crisp	x in R	None	1	x
Fuzzy	μ in [0,1]	0 <= μ <= 1	1	μ
IFT	(μ, ν)	$\mu + \nu$ <= 1	2	μ - ν
SVN Tensor	(T, I, F)	T + I + F <= 3	3	T - (I+F)/2
Plithogenic	$(T, I, F)^{(k)} + C$	C symmetric, diag=0	3n + n(n-1)/2	PTS(A)

Table 6.1. Five tensor classes: entry structure, epistemic parameters per entry, and canonical score function. IFT denotes Intuitionistic Fuzzy Tensor; PTS denotes Plithogenic Truth Score.

Theorem 6.1 (Strict containment hierarchy). *Crisp strictly contained in Fuzzy strictly contained in IFT strictly contained in SVN strictly contained in Plithogenic. Each inclusion is strict; no reverse map is lossless in general.*

Proof. The embeddings are explicit. Crisp into Fuzzy: x maps to μ = x for x in $[0, 1]$; reverse fails for μ in (0, 1) when x must be a real number. Fuzzy into IFT: μ maps to (μ, 0); the IFT (0.6, 0.3) has no fuzzy pre-image. IFT into SVN: (μ, ν) maps to (μ, 1 - μ - ν, ν); the SVN entry (0.8, 0.7, 0.3) with T + I + F = 1.8 > 1 is inadmissible in IFT. SVN into Plithogenic: via n = 1 with C = [0]; an SVN tensor that loses the contradiction matrix C cannot recover it from its image under the projection. ■

Theorem 6.2 (Information monotonicity). *Let I(class) denote the number of independent parameters per entry in each class. Then* $\mathrm{I(Crisp)} = 1 < \mathrm{I(IFT)} = 2 < \mathrm{I(SVN)} = 3 < \mathrm{I(Plitho.)} = 3n + 1$ *for plithogenic tensors with n at least 2 attributes.*

Theorem 6.3 (Aggregation recovery). SVN tensor contraction with weight vector w recovers the SVNWA of Ye (2014): $\mathrm{SVNWA}(t_1, \dots, t_d; w) = (1 - \prod_j (1 - T_j)^{w_j}, \prod_j I_j^{w_j}, \prod_j F_j^{w_j})$.

6.2 Plithogenic Operations

For two plithogenic entries p = (T, I, F) and q = (T', I', F') with contradiction degree c between their attributes, the plithogenic conjunction and disjunction are defined as follows. The plithogenic conjunction is AND_c(p, q) = (T times T' - c times T times (1 - T'), (I + I')/2, F + F' - F times F'), and the plithogenic disjunction is OR_c(p, q) = (T + T' - T times T', (I + I')/2, F times F' - c times F times (1 - F')). The operators AND_c and OR_c generalise the standard neutrosophic conjunction and disjunction by the contradiction factor c in $[0, 1]$. When c = 0 (no contradiction), they reduce to the standard neutrosophic operators.

6.3 The Decision Optimality and Contradiction Visibility Theorems

Theorem 6.4 (Decision Optimality under indeterminacy). *If any expert assigns I > 0 to any criterion for any alternative, there exist instances where the argmax under SVN representation differs from the argmax under IFT representation.*

Proof. We exhibit a witness with m = 2 alternatives, n = 1 criterion, and e = 1 expert. Let A_1 have SVN triplet (T = 0.7, I = 0.5, F = 0.1), yielding $S^{\mathrm{SVN}}(A_1) = T - (I + F)/2 = 0.7 - (0.5 + 0.1)/2 = 0.40$. Let A_2 have IFT pair (μ = 0.6, ν = 0.2), yielding $S^{IFT}(A_2)$ = 0.6 - 0.2 = 0.40. The IFT representation forces A_1 to (μ = 0.7, ν = 0.1) by discarding I = 0.5, yielding $S^{IFT}(A_1)$ = 0.7 - 0.1 = 0.60 and ranking A_1 above A_2. The SVN representation preserves I = 0.5, yielding $S^{SVN}(A_1)$ = 0.40 and $S^{SVN}(A_2)$ = 0.6 - 0.2/2 = 0.50, ranking A_2 above A_1. The decision is reversed. ■

Theorem 6.5 (Contradiction Visibility). *Define the Contradiction Index of alternative i as* $CI(i) = \frac{2}{n(n-1)} \sum_{k<l} C_{kl} \, |T_i^{(k)} - T_i^{(l)}|$. *Then CI(i) = 0 for any SVN encoding (n = 1). For plithogenic representations with n at least 2, CI(i) > 0 whenever* C_{kl} *> 0 and* $|T_i^{(k)}$ - $T_i^{(l)}|$ > *0. The plithogenic score modifies the ranking whenever the margin* $S^{\mathrm{SVN}}(i) - S^{\mathrm{SVN}}(j)$ *is less than* $\alpha(CI(i) - CI(j))$.

6.4 Application: LLM-Based Clinical Decision Support

6.4.1 Problem Setup

We evaluate three language-model-based clinical decision support systems: A_1 (GPT-4o CDSS), A_2 (Claude CDSS), and A_3 (Llama-3 CDSS). The evaluation uses four criteria: C_1 (diagnostic accuracy, weight 0.30), C_2 (uncertainty communication, weight 0.20), C_3 (hallucination risk level, weight 0.25, treated as a cost criterion where lower truth is better), and C_4 (clinical safety, weight 0.25). Three experts are involved: E_1 (emergency physician, weight 0.40), E_2 (AI safety specialist, weight 0.35), and E_3 (medical informaticist, weight 0.25).

6.4.2 Expert-Weighted Aggregated SVN Values

Alt.	**Criterion**	**T**	**I**	**F**
A1	C1	0.877	0.073	0.050
A1	C2	0.552	0.348	0.100

A1	C3	0.175	0.247	0.672
A1	C4	0.778	0.122	0.100
A2	C1	0.775	0.125	0.100
A2	C2	0.875	0.078	0.048
A2	C3	0.067	0.115	0.823
A2	C4	0.902	0.058	0.041
A3	C1	0.625	0.223	0.152
A3	C2	0.525	0.375	0.100
A3	C3	0.223	0.275	0.578
A3	C4	0.677	0.223	0.100

Table 6.2. Expert-weighted SVN aggregates with weights W_E = (0.40, 0.35, 0.25). Note A_2's very low T on C_3 (0.067) combined with high C_4 (0.902) creates a large contradiction gap that contributes to A_2's elevated Contradiction Index.

6.4.3 Aggregated Scores across All Five Tensor Types

Tensor type	A1	A2	A3	Ranking	Gap A2-A1
Crisp	0.3838	0.3945	0.2824	A2 > A1 > A3	0.0108
Fuzzy	0.6119	0.6498	0.5175	A2 > A1 > A3	0.0379
IFT	0.3838	0.3945	0.2824	A2 > A1 > A3	0.0108
SVN	0.4059	0.4740	0.2669	A2 > A1 > A3	0.0681
Plithogenic	0.3143	0.3704	0.2094	A2 > A1 > A3	0.0561

Table 6.3. Aggregated scores across the five tensor types. Crisp and IFT yield identical scores because IFT equals T - F when I is discarded.

The ranking $A_2 > A_1 > A_3$ is consistent across all five tensor representations, but the gap between A_1 and A_2 varies substantially. Under Crisp and IFT, the gap is only 0.0108, suggesting that A_1 and A_2 are nearly tied. Under SVN, the gap widens to 0.0681 because A_1's high truth on diagnostic accuracy is penalised by its high indeterminacy on uncertainty communication. Under the plithogenic representation, the gap settles at 0.0561 after accounting for both the indeterminacy penalty and the contradiction penalty: A_1's high accuracy comes at the cost of internal contradiction with the hallucination risk criterion ($C_{1,3}$ = 0.70 in the contradiction matrix), a tension that the lower-resolution representations do not surface.

6.4.4 The Plithogenic Truth Score and Sensitivity Analysis

The Plithogenic Truth Score is defined as $PTS(A, \alpha) = \text{Truth}_A - \alpha CI(A) - (I + F)_{\text{penalty}}$, where $\text{Truth}_A = \sum_k w_{Ck} T_A^{(k)}$, $(I + F)_{\text{penalty}} = \frac{1}{2} \sum_k w_{Ck} (I_A^{(k)} + F_A^{(k)})$, and $CI(A) = \frac{2}{n(n-1)} \sum_{k<l} C_{kl} |T_A^{(k)} - T_A^{(l)}|$. The parameter α controls the weight of the contradiction penalty; α = 0 corresponds to ignoring contradiction, α = 1 to maximum contradiction penalisation. A sensitivity analysis varying α from 0 to 1 in increments of 0.25 confirms that the ranking $A_2 > A_1 > A_3$ is stable for all α in $[0, 1]$, establishing that the recommendation is robust to uncertainty about the contradiction-penalty weight.

6.5 Plithogenic Tensors versus AHP and TOPSIS

AHP (Analytic Hierarchy Process) and TOPSIS produce a ranking of alternatives. The plithogenic tensor evaluation framework produces a substantially richer report. Across nine dimensions of output (ranking, criterion contradiction report, indeterminacy quantification, score decomposition, ranking stability, expert disagreement, scenario analysis, methodological warning, and explanation of risky alternatives), AHP and TOPSIS provide only the first and partial coverage of the fifth via manual sensitivity analysis. The plithogenic framework provides all nine. The key insight is illustrated by the clinical example: A_1 appears competitive under fuzzy or crisp evaluation because its high truth score on C_1 (diagnostic accuracy = 0.877)

dominates. Only the plithogenic tensor flags that this accuracy comes at the cost of internal contradiction with C_3 (hallucination risk), a tension captured by $C_{1,3}$ = 0.70. When the hallucination risk worsens, crisp and IFT representations miss this signal entirely and incorrectly prefer A_1.

6.6 Methodological Warning: Contradiction Matrix Interpretability

The contradiction matrix C is a powerful structural object but its entries are not automatically interpretable. A high C_{kl} value means that optimising criterion k and criterion l simultaneously is structurally difficult, but this must be validated rather than assumed. The general methodology requires five validation steps. First, expert elicitation: the contradiction matrix should be elicited from domain experts using structured methods such as AHP-style pairwise comparisons, Delphi rounds, or systematic literature review, with at least two independent expert groups providing estimates. Second, cross-expert agreement: inter-rater reliability (intraclass correlation or Krippendorff's α) should be computed for each elicited C_{kl} value; high disagreement signals that the contradiction is context-dependent. Third, sensitivity analysis: each C_{kl} entry should be varied by plus-or-minus 0.15 and the stability of the ranking checked. Fourth, ablation study: the evaluation should be run with C = 0 (no contradictions) and compared with the full model. Fifth, counterfactual scenarios: the minimum perturbation δ-C that changes the decision should be reported, with a large δ-C indicating a robust recommendation and a small one signalling fragility.

6.7 Conclusion

Chapter 6 has established a rigorous five-level tensor hierarchy with strict containment and monotonically increasing epistemic information capacity. Two novel theorems ground the decision-theoretic consequences. The Decision Optimality theorem proves that SVN tensors can reverse IFT rankings when I > 0; the Contradiction Visibility theorem proves that plithogenic tensors expose inter-attribute tensions that change decisions in safety-critical settings. Applied to language-model-based clinical decision support, the framework produces richer reports than AHP or TOPSIS: ranking, contradiction decomposition, sensitivity analysis, expert disagreement metrics, and scenario analysis. The plithogenic tensor correctly identifies A_1's internal contradiction between diagnostic accuracy and hallucination risk and adjusts the recommendation accordingly. Classical tensor methods learn structure; plithogenic tensors explain tension. Part III

of this book extends the framework with paraconsistent logic, providing the inferential apparatus for reasoning under genuine ontological contradiction.

PART III

Neutrosophic Paraconsistent Logic

Chapter 7

Neutrosophic Paraconsistent Logic: A Formal System

Classical logic enforces the principle of explosion: from any contradiction every proposition is derivable, rendering the system trivial. Da Costa demonstrated in 1963 that explosion is a design choice rather than a logical necessity, and his hierarchy of annotated paraconsistent systems tolerates contradictions while preserving inferential capacity. Smarandache independently introduced neutrosophy in 1998, adding a genuine indeterminacy component alongside truth and falsehood. Despite their complementary strengths, neither system alone captures a distinction that is both philosophically and practically critical: the difference between a contradiction that is ontological, structural, and independent of available information, and one that is epistemic, arising from incomplete information and in principle resolvable. This chapter introduces Neutrosophic Paraconsistent Logic (NPL), a hybrid formal system that strictly extends both predecessors and formalises the ontological-versus-epistemic distinction through a threshold parameter.

7.1 Predecessors and the Gap They Leave

Classical propositional logic obeys the principle of non-contradiction and the law of explosion: from the pair (φ, not-φ) every formula ψ is derivable. This makes classical logic unsuitable for inconsistency-tolerant reasoning. Any theory containing a single contradiction is rendered trivial, since every proposition becomes provable.

Da Costa's annotated paraconsistent logic (LPA), introduced in 1963 and developed through the hierarchy C_1, C_2, ..., C_omega, replaces the classical principles with a paraconsistent alternative. In LPA, propositions carry annotations (μ, λ) in $[0,1]^2$ that represent degrees of favourable and contrary evidence. A proposition is paraconsistent when $\mu + \lambda > 1$ (both evidence streams are strong). The system collapses to trivial only when $\mu = \lambda = 1$, the LPA-trivial state. LPA's limitation, from the perspective of the present book, is that it has no mechanism for genuine indeterminacy I independent of μ and λ. Two contradictions with identical (μ, λ) but structurally different origins, one arising from incomplete information and one from structural incompatibility, receive identical representations in LPA.

Smarandache's single-valued neutrosophic logic (SVNL) carries propositions with triplets (T, I, F) subject to the constraint T + I + F = 1 (normalisation on the simplex). This enables representing indeterminate propositions but prevents full independence: knowing any two components determines the third. The normalisation constraint conflates the three epistemic dimensions. A proposition with high T and high F (genuine contradiction) must have low I, which is counter-intuitive in cases where contradiction and indeterminacy coexist: for instance, a proposition that is simultaneously well-evidenced on both sides (high μ and λ) and genuinely indeterminate in its structural resolution (high I).

Weber's dialethic mathematics, published in 2022, provides the most systematic recent development of dialethism, arguing that every true and complete theory necessarily contains contradictions. Beall, reviewing Weber the same year, raises a fundamental asymmetry objection: the dialethic programme accepts glutty logical possibilities (propositions that are both true and false) while systematically excluding gappy possibilities (propositions that are neither true nor false). NPL addresses both Weber's positive thesis and Beall's critique directly. Weber's ontological gluts correspond to NPL-Para-onto (propositions with $\mu + \lambda > 1$ and $I > \theta$). Beall's gaps correspond to NPL-PC (propositions with low μ and λ). NPL treats both symmetrically within the same cubic space $[0,1]^3$, with no categorical asymmetry between them.

7.2 The NPL Formal System

7.2.1 Syntax and Annotations

Definition 7.1 (NPL proposition). *An NPL proposition is a pair $\varphi:(\mu, \lambda, I)$ where φ is a propositional formula and (μ, λ, I) in $[0,1]^3$. The components are: μ in $[0, 1]$, the degree of favourable evidence; λ in $[0, 1]$, the degree of contrary evidence; and I in $[0, 1]$, the degree of genuine indeterminacy. No normalisation constraint is imposed: μ, λ, I vary independently.*

7.2.2 Logical Connectives

Definition 7.2 (NPL connectives). *For propositions $\varphi:(\mu_1, \lambda_1, I_1)$ and $\psi:(\mu_2, \lambda_2, I_2)$, the NPL connectives are: negation, $\neg\varphi : (I_1, \lambda_1, \mu_1)$; conjunction, $\varphi \wedge \psi : (\min(\mu_1, \mu_2), \max(\lambda_1, \lambda_2), \max(I_1, I_2))$; disjunction, $\varphi \vee \psi : (\max(\mu_1, \mu_2), \min(\lambda_1, \lambda_2), \max(I_1, I_2))$; implication, $\varphi \rightarrow \psi : (\min(1, 1 - \mu_1 + \mu_2), \max(0, \lambda_2 - \lambda_1), \max(0, I_2 - I_1))$.*

The negation rule swaps favourable and contrary evidence while preserving indeterminacy, consistent with the intuition that negating a proposition does not reduce its ontological indeterminacy. The conjunction rule propagates the maximum indeterminacy, ensuring that combining with an ontologically indeterminate proposition does not artificially reduce overall indeterminacy. The implication rule follows the Lukasiewicz-style construction adapted to the triadic setting.

7.2.3 Epistemic States

Definition 7.3 (Epistemic classification). *Given a threshold θ in (0, 1) with default $\theta = 0.5$, an NPL proposition φ:(μ, λ, I) is classified as: NPL-T if $\mu = \lambda = I = 1$ (trivial, the unique collapse state); NPL-Para-onto if $\mu + \lambda > 1$ and $I > \theta$ (ontological contradiction); NPL-Para-epist if $\mu + \lambda > 1$ and I at most θ (epistemic contradiction); NPL-I if $I > \theta$ and $\mu + \lambda$ at most 1 (pure indeterminacy); NPL-V if $\mu > 0.5$ and $\mu > \lambda$ (predominantly true); NPL-F if $\lambda > 0.5$ and $\lambda > \mu$ (predominantly false); NPL-PC otherwise (paracomplete / uncertain).*

7.2.4 Inference Rules

Definition 7.4 (Inference system H-NPL). *The system H-NPL contains four rules. NPL-MP (modus ponens): from φ:(mu_1, $lambda_1$, I_1) and (φ implies ψ):(mu_2, $lambda_2$, I_2), conclude ψ:(min(mu_1, mu_2), max($lambda_1$, $lambda_2$), max(I_1, I_2)). NPL-IC (I-conservation): if φ is NPL-Para-onto, then for any ψ derived from φ, I(ψ) at least I(φ). NPL-andI (conjunction introduction) and NPL-andE (conjunction elimination) follow the standard rules with the connectives of Definition 7.2.*

7.3 The Eight Central Theorems

Theorem 7.1 (Proper extension of LPA). *NPL strictly extends LPA: every LPA proposition (μ, λ) maps to NPL as (μ, λ, 0), but NPL admits propositions with $I > 0$ that have no LPA counterpart. Furthermore, the LPA-trivial state (1, 1) maps to NPL as (1, 1, 0), which is classified as NPL-Para-epist and is therefore not trivial in NPL and does not permit arbitrary derivation.*

Proof. The mapping $phi_L PA$: (μ, λ) maps to $phi_N PL$: (μ, λ, 0) is injective and preserves all LPA connectives. The LPA-trivial state (1, 1) maps to (1, 1, 0). By Definition 7.3, NPL-T

requires $\mu = \lambda = I = 1$; since I = 0 is not 1, (1, 1, 0) is classified as NPL-Para-epist (since I = 0 at most θ). It is not trivial and does not permit arbitrary derivation. Any NPL proposition with I > 0 has no pre-image under the LPA-to-NPL map, confirming strictness. ■

Theorem 7.2 (Proper extension of SVNL). *NPL strictly extends SVNL: every normalised SVNL triplet* (T, I, F) *with T + I + F = 1 maps injectively into NPL as (T, F, I), but NPL admits propositions with* $\mu + \lambda + I$ *!= 1 that have no SVNL counterpart.*

Proof. Define the embedding e : (T, I, F) maps to (T, F, I) in $[0,1]^3$. Since T + I + F = 1, the image of e lies on the two-simplex, a proper subset of $[0,1]^3$. Any NPL proposition with $\mu + \lambda + I$ different from 1 falls outside this simplex and has no pre-image under e, establishing strictness. Connective preservation follows by direct verification against Definition 7.2. ■

Theorem 7.3 (Failure of explosion). *For any NPL proposition φ that is not NPL-T, the set {φ, not-φ} does not derive an arbitrary proposition ψ:(1, 0, 0) under H-NPL.*

Proof. Suppose φ:(μ, λ, I) is not NPL-T, so (μ, λ, I) is not equal to (1, 1, 1). Then not-φ:(λ, μ, I). To derive ψ:(1, 0, 0) via NPL-MP we need φ implies ψ : (1, 0, I). By Definition 7.2, the μ-component of φ implies ψ is min(1, 1 - μ + 1) = 1 only if μ = 1. If μ < 1, then φ implies ψ has μ-component less than 1, and NPL-MP yields a conclusion with μ-component less than 1, which is not ψ:(1, 0, 0). The only state from which arbitrary derivation is possible is NPL-T. ■

Theorem 7.4 (Distinguishability of contradiction types). *Two propositions* phi_1*:(μ, λ,* I_1*) and* phi_2*:(μ, λ,* I_2*) with the same (μ, λ) but* $I_1 > \theta > I_2$ *produce structurally different inference trees under H-NPL.*

Proof. Both phi_1 and phi_2 are NPL-Para since $\mu + \lambda > 1$. By Definition 7.3, phi_1 is NPL-Para-onto and phi_2 is NPL-Para-epist. By the I-conservation rule, any conclusion derived from phi_1 satisfies I(conclusion) at least $I_1 > \theta$, classifying it as ontologically indeterminate. No such constraint applies to conclusions from phi_2. The inference trees are therefore structurally distinct. Classical logic, LPA, and SVNL cannot make this distinction for the same (μ, λ) values. ■

Theorem 7.5 (Monotonicity of I under I-conservation). *If φ is NPL-Para-onto with I(φ) = $I_0 > \theta$, then for any chain of derivations psi_1, psi_2, ..., psi_n under H-NPL where each step applies NPL-IC, I(psi_k) at least I_0 for all k.*

Proof. By induction. Base case: NPL-IC applied to φ yields psi_1 with I(psi_1) at least I_0. Inductive step: if I(psi_k) at least $I_0 > \theta$, then psi_k is itself NPL-Para-onto (since $\mu + \lambda > 1$ is preserved by NPL-MP's min and max structure when φ is ontologically contradictory). NPL-IC applied to psi_k yields psi_{k+1} with I(psi_{k+1}) at least I(psi_k) at least I_0. The sequence I(psi_k) is monotonically non-decreasing and bounded above by 1. ■

Theorem 7.6 (NPL generalises classical logic). *Classical logic is a limiting case of NPL: a proposition φ:(μ, λ, I) with I = 0, μ in {0, 1}, and λ = 1 - μ behaves identically to its classical counterpart.*

Proof. When I = 0 and μ = 1, λ = 0: φ is NPL-V with maximum favourable evidence and no contrary evidence. Negation yields not-φ:(0, 1, 0) = NPL-F. Conjunction and disjunction reduce to min and max over {0, 1}, equivalent to classical AND and OR. Explosion fails only at NPL-T; with I = 0, NPL-T requires $\mu = \lambda$ = I = 1, which cannot arise from classical assignments. The classical tautology φ or not-φ maps to (max(1, 0), min(0, 1), 0) = (1, 0, 0) = NPL-V. ■

Theorem 7.7 (Conservation of ontological contradictions). If φ is NPL-Para-onto and ψ is any proposition with μ(ψ) > 0, then (φ and ψ) is NPL-Para-onto. Proof. Let φ:(μ_1, λ_1, I_1) with $\mu_1 + I_1 > 1$ and $I_1 > \mu_1$, and let ψ:(μ_2, λ_2, I_2) with $\mu_2 > 0$. By Definition 7.2, "φ and ψ" has triplet (min(μ_1, μ_2), max(λ_1, λ_2), max(I_1, I_2)). Since $I_1 > 1 - \mu_1 \geq 1 - \min(\mu_1, \mu_2)$, it follows that $\max(I_1, I_2) \geq I_1 > 1 - \min(\mu_1, \mu_2)$. Hence $\min(\mu_1, \mu_2) + \max(I_1, I_2) > 1$, confirming NPL-Para. Moreover $\max(I_1, I_2) \geq I_1 > \mu_1 \geq \min(\mu_1, \mu_2)$, confirming NPL-Para-onto.. ■

Theorem 7.8 (Mutual independence of μ, λ, I). *There exist no logical dependencies among μ, λ, I in NPL: for any triplet (mu_0, $lambda_0$, I_0) in $[0,1]^3$, there exists a coherent NPL proposition with that exact annotation.*

Proof. By construction: any tuple (name, mu_0, $lambda_0$, I_0) with (mu_0, $lambda_0$, I_0) in $[0,1]^3$ is a well-formed NPL proposition. No axiom or inference rule constrains the joint distribution of the three components, in contrast to SVNL where $\mu + \lambda$ + I = 1 holds.

Independence is confirmed by a ten-point stress test that includes (1, 1, 1) = NPL-T, (1, 1, 0) = NPL-Para-epist, (0, 0, 0.9) = NPL-I, and (0.9, 0.1, 0.05) = NPL-V, all coherent and distinct. ■

7.4 Empirical Validation in Three Domains

We validate NPL on three domains where the ontological-versus-epistemic distinction has concrete implications. Annotations follow a consistent protocol: μ = degree of documented favourable evidence, λ = degree of documented contrary evidence, I = structural irreducibility of the tension (high I if the tension persists independently of additional information).

7.4.1 Indigenous Rights in Ecuador: Colonial Contradiction

Ecuador's 2008 Constitution recognises in Articles 56 through 60 the collective rights of indigenous nationalities and in Articles 71 through 74 the rights of nature. Simultaneously, the state pursues systematic extractive policies (petroleum contracts, mining in ancestral territories, the Mirador project) that violate those rights. This is not a case of insufficient legal clarity: the contradiction is structural to the plurinational extractive state model itself.

Proposition	μ	λ	**I**	**State**	**Type**
Collective rights recognised	0.92	0.15	0.20	NPL-V	-
Extractive policy active	0.88	0.12	0.18	NPL-V	-
Compatibility of rights and extraction	0.22	0.85	0.78	NPL-Para	Ontological
Prior consultation effective	0.30	0.75	0.60	NPL-Para	Ontological
State sovereignty	0.80	0.78	0.85	NPL-Para	Ontological

over resources					
Judicial remedy effective	0.35	0.70	0.55	NPL-Para	Ontological

Table 7.1. NPL annotations for the Indigenous Rights case ($\theta = 0.5$).

Four of six propositions are ontologically contradictory (I > 0.5). The highest productivity score belongs to State sovereignty over resources (I = 0.85): the tension between two sovereignties, national and indigenous territorial, is structural to the model, not resolvable by more information or better enforcement mechanisms alone. LPA would represent all four as identically LPA-paraconsistent; NPL distinguishes the structural from the resolvable.

7.4.2 Cross-Cultural AI Alignment

A value-alignment system must simultaneously accommodate individual autonomy (dominant in liberal Western frameworks) and communal decision-making (prevalent in Andean and ubuntu frameworks). This tension is not resolvable by aggregating more training data: it arises from structurally different ontologies of personhood, relational versus individualist, not from incomplete representation.

Value proposition	μ	λ	**I**	**State**	**Type**
Individual autonomy priority	0.90	0.72	0.80	NPL-Para	Ontological
Communal decision priority	0.88	0.75	0.82	NPL-Para	Ontological
Individual data privacy	0.85	0.60	0.55	NPL-Para	Ontological

Common good over individual	0.82	0.70	0.75	NPL-Para	Ontological
AI can be culturally neutral	0.20	0.90	0.35	NPL-Para	Epistemic
Coexistence of incompatible values	0.75	0.65	0.90	NPL-Para	Ontological

Table 7.2. NPL annotations for the AI alignment case (θ = 0.5).

AI can be culturally neutral is the only epistemic contradiction (I = 0.35 < θ): the empirical evidence strongly disfavours neutrality, but this tension could in principle be reduced with better cross-cultural training data. All other value tensions are ontological: structural incompatibilities between world-ontologies that no amount of data will eliminate. Pluriversal AI alignment requires systems capable of NPL-Para-onto representation, holding both values simultaneously without forced resolution.

7.4.3 Wave-Particle Duality in Quantum Mechanics

The wave-particle duality provides the paradigmatic case of experimentally verified ontological contradiction. The photon exhibits wave properties (double-slit interference) and particle properties (photoelectric effect) depending on the measurement apparatus, not on the observer's ignorance. Heisenberg's uncertainty principle is ontological, not epistemic: the uncertainty is not due to measurement limitations but to the structure of quantum reality itself.

Proposition	μ	λ	**I**	**State**	**Type**
Photon has wave properties	0.97	0.80	0.92	NPL-Para	Ontological
Photon has particle	0.95	0.82	0.90	NPL-Para	Ontological

properties					
Bohr complementarity	0.80	0.45	0.65	NPL-Para	Ontological
Local hidden variables resolve	0.25	0.88	0.30	NPL-Para	Epistemic
Uncertainty is ontological	0.78	0.55	0.70	NPL-Para	Ontological

Table 7.3. NPL annotations for the wave-particle duality case (θ = 0.5).

7.5 Comparative Analysis

Table 7.4 demonstrates the discriminative power of NPL against classical logic, LPA, and SVNL on a representative set of propositions. The first two rows are critical: identical (μ, λ) = (0.85, 0.80) but different I. Classical logic, LPA, and SVNL return identical classifications; NPL distinguishes them structurally.

Proposition	**CL**	**LPA**	**SVNL**	**NPL**
Onto. contradiction (0.85, 0.80, 0.90)	TRUE	LPA-Para	INVALID	NPL-Para-onto
Epist. contradiction (0.85, 0.80, 0.10)	TRUE	LPA-Para	INVALID	NPL-Para-epist
Clearly true (0.90, 0.10, 0.05)	TRUE	LPA-True	SVNL-True	NPL-V
Clearly false (0.10, 0.90, 0.05)	FALSE	LPA-False	SVNL-False	NPL-F
Max uncertainty	TRUE	LPA-Para	INVALID	NPL-PC

(0.50, 0.50, 0.50)				
NPL-trivial state (1, 1, 1)	TRUE	LPA-Trivial	INVALID	NPL-T
LPA-trivial in NPL (1, 1, 0)	TRUE	LPA-Trivial	INVALID	NPL-Para-epist
Pure indeterminacy (0.30, 0.30, 0.80)	FALSE	LPA-Paracomp.	SVNL-Indet.	NPL-I
Wave-particle (0.97, 0.80, 0.92)	TRUE	LPA-Para	INVALID	NPL-Para-onto

Table 7.4. System comparison: same data, different representational power. INVALID indicates that $\mu + \lambda + I$ differs from 1, making SVNL unable to represent the proposition.

Row 7 is particularly significant: the LPA-trivial state (1, 1) maps to NPL as (1, 1, 0), which NPL classifies as NPL-Para-epist, a manageable paraconsistent state, not a system collapse. NPL is strictly more tolerant than LPA at the critical edge case where LPA explodes.

7.6 NPL versus LP, Belnap's FOUR, and Relevance Logics

7.6.1 Priest's Logic of Paradox

Priest's LP, introduced in 1979, is a three-valued paraconsistent logic with values {T, F, B} where B (both) designates propositions that are simultaneously true and false. Explosion fails in LP because B is designated. LP's limitations relative to NPL are threefold. First, LP is non-graded: a proposition is either B or not, with no measure of how strongly both values are supported. NPL encodes this as $\mu + \lambda$, which can range continuously from 0 to 2. Second, LP has no indeterminacy dimension I independent of the true and false components. Third, and most critically, LP cannot distinguish ontological from epistemic contradiction. Two propositions, one arising from genuine structural incompatibility and one from incomplete information, receive identical representations as B in LP.

Proposition 7.1 (LP embeds in NPL). *Every LP proposition maps injectively into NPL as follows: T maps to (1.0, 0.0, 0.0) = NPL-V; F maps to (0.0, 1.0, 0.0) = NPL-F; B maps to*

(1.0, 1.0, 0.0) = NPL-Para-epist. The embedding is injective but not surjective: NPL admits triplets (μ, λ, I) with $0 < \mu, \lambda < 1$ and $I > 0$ that have no LP counterpart.

7.6.2 Belnap's Four-Valued Logic

Belnap's FOUR, introduced in 1977, has four values: T (told true only), F (told false only), B (told both), and N (told neither). FOUR is more expressive than LP because it adds N alongside B. However, FOUR remains non-graded: both B and N are binary categories. Two contradictions of radically different strength, one with $\mu = 0.55$ and $\lambda = 0.51$ and one with $\mu = 0.95$ and $\lambda = 0.90$, map identically to B in FOUR. NPL differentiates them by contradiction degree $\max(0, \mu + \lambda - 1)$: 0.06 versus 0.85 respectively. More critically, FOUR has no mechanism to distinguish within B between contradictions that are structurally irreducible (high I) and those that are information-sensitive (low I).

Proposition 7.2 (FOUR embeds in NPL). *Every FOUR value maps injectively into NPL: T maps to (1.0, 0.0, 0.0) = NPL-V; F maps to (0.0, 1.0, 0.0) = NPL-F; B maps to (1.0, 1.0, I) for any I in* $[0, 1]$*, which is NPL-Para-epist or NPL-Para-onto depending on I; N maps to (0.0, 0.0, 0.0) = NPL-PC.*

7.6.3 Relevance Logics

Relevance logics, developed by Anderson and Belnap in 1975, avoid explosion not by admitting contradictions as tolerable but by restricting implication: A implies B is valid only when A is relevant to B. Relevance logics and NPL address the explosion problem through fundamentally different strategies. Relevance logics restrict the implication connective; NPL keeps a modified implication and instead restricts the trivial state to a single point. Relevance logics do not provide a representational framework for degrees of evidence or indeterminacy; they are proof-theoretic systems, not annotation-based. They cannot express that a proposition has $\mu = 0.85$ favourable evidence, $\lambda = 0.80$ contrary evidence, and $I = 0.90$ structural indeterminacy. NPL can.

7.6.4 The Russell Set as NPL-T

Weber's central example, the Russell set R = {x : x not in x}, which is both a member of itself and not a member of itself, provides an instructive test case. Under unrestricted comprehension, R satisfies R in R ($\mu = 1$) and R not in R ($\lambda = 1$). The contradiction is not epistemic: no additional set-theoretic information will resolve it; it is constitutive of the definition itself ($I = 1$). The

Russell set therefore maps to NPL annotation (1, 1, 1) = NPL-T, the unique trivial state. This is not a deficiency of NPL but its correct formal behaviour: NPL-T is the sole state from which arbitrary derivation is possible, corresponding precisely to the explosion that Russell's paradox triggers in naive set theory. What NPL adds is that NPL-T is a single point in $[0,1]^3$, not an entire category. All other paraconsistent propositions, including Weber's glutty arithmetic numbers and topological borderline cases, map to NPL-Para-onto or NPL-Para-epist and are inferentially controlled.

7.7 Discussion

The central formal contribution of NPL, the threshold-based distinction between ontological and epistemic contradiction, has direct philosophical consequences. An epistemic contradiction (I at most θ) calls for more investigation: gather data, clarify concepts, run experiments. An ontological contradiction (I > θ) calls for a different response: design institutions, pedagogies, and inference systems capable of inhabiting the tension productively rather than forcing its resolution. This distinction underlies the broader project of a rationality adequate for subjects who exist in structural, irreducible contradiction: colonial subjects, culturally hybrid knowers, quantum physicists confronting complementarity. NPL provides the formal infrastructure.

Three limitations deserve acknowledgement. First, the threshold θ is a free parameter; its setting determines the ontological-versus-epistemic boundary and should ideally be calibrated from domain-specific evidence rather than set to 0.5 by default. Second, the I annotations in our empirical cases are expert-assigned degrees requiring validation against independent measurement procedures. Third, the NPL inference rules constitute a Hilbert-style system whose completeness with respect to the three-valued semantics remains an open formal question.

7.8 Conclusion

Neutrosophic Paraconsistent Logic offers three advances over its predecessors. First, it strictly extends both LPA and SVNL, recovering each as a special case while admitting propositions neither can represent. Second, it introduces the ontological-versus-epistemic distinction via a threshold parameter, enabling inference systems to treat structurally irreducible contradictions differently from information-sensitive ones. Third, the I-conservation rule ensures that ontological indeterminacy propagates through derivation chains, preventing artificial reduction of genuine uncertainty. Chapter 8 establishes the connection between NPL and the Type-k

Neutrosophic Sets of Chapter 3, mapping the seven epistemic states of NPL to specific configurations in the Type-k framework.

Chapter 8

NPL and Type-k: Connections and Alignments

This chapter establishes the formal connection between Neutrosophic Paraconsistent Logic (Chapter 7) and Type-k Neutrosophic Sets (Chapter 3). The two frameworks were developed for different purposes: NPL provides an inferential apparatus for reasoning under contradiction, while Type-k provides a representational apparatus for nested triadic uncertainty. We show that the seven epistemic states of NPL correspond to specific configurations in the Type-2 framework, that the ontological-versus-epistemic distinction of NPL maps to a structural property of the I-sub-triplet at Type-2 depth, and that the I-conservation rule of NPL has a natural Type-k analogue. The chapter closes with the NPL-Type-k alignment proposition, which formalises the operational compatibility of the two frameworks.

8.1 Motivation: Two Frameworks, One Goal

NPL and Type-k both extend the classical neutrosophic framework, but along orthogonal axes. NPL extends the framework in the inferential dimension: it equips the triadic representation with inference rules that block explosion and preserve ontological indeterminacy through derivation chains. Type-k extends the framework in the representational dimension: it nests the triadic representation within itself, so that each component of a Type-1 triplet is itself characterised by a triplet at the next level. The two extensions are independent: an NPL inference system can operate on Type-1 propositions (as in Chapter 7) or on Type-k propositions (as we develop here), and a Type-k representation can be evaluated under classical or paraconsistent inference rules.

The combination of the two extensions is more powerful than either alone. NPL provides the formal language for distinguishing ontological from epistemic contradiction; Type-k provides the formal language for representing meta-uncertainty about each component of an evaluation. When the two are combined, the meta-uncertainty about the indeterminacy component of a Type-1 evaluation can itself be classified as ontological or epistemic by reference to the indeterminacy of the indeterminacy at Type-2 depth. This nested classification is operationally relevant for the auditing of language models, which often produce evaluations whose indeterminacy is itself uncertain in a structural rather than merely informational sense.

8.2 Mapping NPL States to Type-2 Configurations

Recall from Chapter 3 that a Type-2 neutrosophic representation of a proposition is a triplet of sub-triplets: $((T_T, I_T, F_T), (T_I, I_I, F_I), (T_F, I_F, F_F))$. Recall from Chapter 7 that an NPL proposition is a triplet (μ, λ, I) classified into one of seven epistemic states by the threshold θ. The natural mapping between the two frameworks proceeds in two steps. First, the NPL components (μ, λ, I) correspond to the Type-1 components (T, F, I) of the underlying proposition, under the identification (μ, λ, I) = (T, F, I). Second, the structural type of the NPL contradiction (ontological versus epistemic) corresponds to the I-sub-triplet of the Type-2 representation: an ontological contradiction has I_I close to zero (we are certain that the indeterminacy is real), while an epistemic contradiction has I_I close to one (we are uncertain whether the indeterminacy reflects genuine ignorance or merely incomplete information).

Definition 8.1 (NPL-Type-2 alignment). *Let φ be an NPL proposition with annotation (μ, λ, I). The Type-2 alignment of φ is the Type-2 element ($(\mu, 0, 1 - \mu), (T_I, I_I, F_I), (\lambda, 0, 1 - \lambda)$), where the I-sub-triplet (T_I, I_I, F_I) is determined by the epistemic classification of φ: if φ is NPL-Para-onto then (T_I, I_I, F_I) = (I, 0, 1 - I) (certain ontological indeterminacy); if φ is NPL-Para-epist then (T_I, I_I, F_I) = (I, 1 - I, 0) (uncertain epistemic indeterminacy); for other classifications, the I-sub-triplet is set to (I, 0.5, 0.5 - I/2) as a neutral default.*

Proposition 8.1 (NPL-Type-k alignment). *Under the alignment of Definition 8.1, the scalar projection of the Type-2 representation recovers the original NPL triplet, and the seven epistemic states of NPL correspond to seven distinct regions in the Type-2 space $NS^{(2)}(U)$.*

Proof. The scalar projection of a Type-2 element maps each sub-triplet to its T-component, yielding (μ, T_I, λ). The T-component of the I-sub-triplet is the original I value by construction. The projection therefore recovers (μ, λ, I) up to a coordinate permutation. The seven NPL states correspond to seven distinct regions in Type-2 space because the alignment of Definition 8.1 assigns distinct I-sub-triplets to NPL-Para-onto and NPL-Para-epist, distinguishing them at Type-2 depth in a way that the Type-1 representation cannot. ■

8.3 Ontological versus Epistemic Contradiction at Type-2 Depth

The central conceptual gain of the NPL-Type-k alignment is that the ontological-versus-epistemic distinction, which in NPL is captured by a threshold on a scalar I, has a Type-2 representation as

a structural property of the I-sub-triplet. Specifically, an NPL-Para-onto proposition has an I-sub-triplet of the form (I, 0, 1 - I): the truth of the indeterminacy is high (equal to I), the indeterminacy of the indeterminacy is zero (we are certain about the indeterminacy), and the falsity of the indeterminacy is the complement of I (the indeterminacy is not falsified). This configuration encodes the philosophical commitment of NPL-Para-onto: the contradiction is structurally real and our knowledge of its structural nature is itself reliable.

An NPL-Para-epist proposition has an I-sub-triplet of the form (I, 1 - I, 0): the truth of the indeterminacy is again I, but now the indeterminacy of the indeterminacy is 1 - I, signalling that we are uncertain about whether the indeterminacy is real or merely apparent. This configuration encodes the alternative philosophical commitment: the contradiction may dissolve under further investigation, and our current assessment of its structural nature is itself uncertain. The Type-2 representation makes this distinction explicit at the level of the data, not merely at the level of the inference rules.

Within the inference system, the I-conservation rule of NPL has a natural Type-k extension. We state it for Type-2 and remark that the extension to general k is immediate.

> **Definition 8.2 (Type-2 I-conservation).** *If φ is a Type-2 NPL proposition with I-sub-triplet (I, 0, 1 - I) (corresponding to NPL-Para-onto), then for any proposition ψ derived from φ under the Type-2 inference rules, the I-sub-triplet of ψ satisfies $T_I(\psi)$ at least $T_I(\varphi)$ and $I_I(\psi)$ at most $I_I(\varphi)$. That is, both the truth of the indeterminacy and the certainty of that truth are preserved across derivations.*

8.4 Implications for Multi-Vendor Auditing

The NPL-Type-k alignment has direct implications for the multi-vendor auditing protocols of Chapter 5 and the unified epistemic auditing protocol of Chapter 11. When a panel of language models is asked to evaluate a contested proposition, the variation across models in the truth, indeterminacy, and falsity components can be analysed at two levels. At Type-1 depth, the variation is captured by the spread of the (T, I, F) triplets across vendors. At Type-2 depth, the variation is captured by the spread of the nine sub-component values, including in particular the spread of the I_I sub-component, which measures the meta-disagreement among models about whether the indeterminacy of the underlying proposition is structural or informational.

Empirically, we expect Type-2 disagreement to be highest precisely on the propositions where NPL would assign NPL-Para-onto: ethical contradictions with no agreed resolution, paradoxes of self-reference, statements involving cross-cultural value tensions, and quantum-mechanical claims that touch on the foundations of the theory. The cross-vendor empirical studies reported in Chapters 5 and 9, which document hyper-truth rates ranging from 62 percent (Mistral) to 100 percent (Alibaba, Anthropic), can be reinterpreted within the NPL-Type-k framework as measurements of the ontological-versus-epistemic profile of contemporary foundation models on canonical paraconsistent phenomena.

8.5 Conclusion

Chapter 8 has established the formal connection between Neutrosophic Paraconsistent Logic and Type-k Neutrosophic Sets. The seven epistemic states of NPL correspond to seven distinct regions in Type-2 space, and the ontological-versus-epistemic distinction of NPL has a structural representation as a property of the I-sub-triplet at Type-2 depth. The I-conservation rule of NPL extends naturally to a Type-k rule that preserves both the truth of the indeterminacy and the certainty of that truth across derivations. The combined NPL-Type-k framework provides both the representational and the inferential apparatus required for the empirical chapters of Part IV, which apply the framework to the auditing of large language models in three complementary settings: hyper-truth elicitation (Chapter 9), hallucination detection beyond softmax (Chapter 10), and unified epistemic auditing (Chapter 11).

PART IV

Epistemic Auditing of Large Language Models

Chapter 9

Breaking the Chains of Probability

The deployment of large language models in high-stakes decision domains has made the robust quantification of epistemic uncertainty a first-order engineering requirement. The dominant architecture of these models, however, is rooted in probability theory, where outcome probabilities are constrained to sum to unity by softmax normalisation. This forces a zero-sum game in which any increase in uncertainty must subtract from truth or falsity. The constraint hinders the ability of language models to distinguish between aleatoric uncertainty (statistical uncertainty inherent in the data) and epistemic uncertainty (model uncertainty due to lack of knowledge), and in particular between not knowing (ignorance) and knowing of a conflict (paradox or contradiction). This chapter presents an empirical investigation that releases language models from the softmax constraint and documents the resulting class of declared epistemic states that probabilistic prompting cannot represent by construction.

9.1 The Probabilistic Constraint and the Collapse of Uncertainty

Recent work on uncertainty quantification for language models has explored several alternatives, including semantic entropy with linguistic invariances (Kuhn, Gal, and Farquhar, 2023), self-consistency checks via SelfCheckGPT (Manakul, Liusie, and Gales, 2023), and conformal abstention policies (Yadkori et al., 2024). These approaches address calibration and abstention but operate within probabilistic representations and inherit their structural limitations.

Neutrosophic logic offers an alternative semantic foundation. It generalises fuzzy and intuitionistic-fuzzy logics by introducing three independent components, truth (T), indeterminacy (I), and falsity (F), each a real number in $[0, 1]$, without the constraint that they sum to unity. This freedom allows the simultaneous expression of high truth, high falsity, and high indeterminacy, a state we call hyper-truth ($T + I + F > 1$). We hypothesise that under unconstrained neutrosophic prompting, current language models will declare hyper-truth at non-trivial rates specifically in cases of paradox and ethical contradiction, while probabilistic prompting will not. The remainder of this chapter tests this hypothesis empirically and frames it within a formal neutrosophic apparatus.

9.2 Formal Preliminaries

Definition 9.1 (Single-valued neutrosophic set). *Let X be a universe of discourse. A single-valued neutrosophic set A on X is the set of ordered quadruples A = {(x,* $T_A(x)$*,* $I_A(x)$*,* $F_A(x)$*) : x in X}, where* $T_A(x)$*,* $I_A(x)$*,* $F_A(x)$ *denote, respectively, the truth-membership degree, the indeterminacy-membership degree, and the falsity-membership degree of x in A. Each of the three functions maps X to the unit interval* $[0, 1]$*, and no constraint is imposed on their sum, which therefore lies in [0, 3].*

Definition 9.2 (Neutrosophic evaluation of a statement). *Given a statement s and an evaluator E, the neutrosophic evaluation of s by E is the ordered triple* $n_E(s) = (T_E(s), I_E(s), F_E(s))$ *in* $[0,1]^3$*, where* $T_E(s)$*,* $I_E(s)$*, and* $F_E(s)$ *denote, respectively, the truth degree, indeterminacy degree, and falsity degree assigned by E to s.*

Definition 9.3 (Hyper-truth). *A neutrosophic evaluation n(s) =* (T, I, F) *in* $[0,1]^3$ *exhibits hyper-truth if and only if* $T + I + F > 1$*. The hyper-truth region is the subset H =* $\{(T, I, F) in [0,1]^3 : T + I + F > 1\}$*.*

Definition 9.4 (Strategy mappings). Each prompting strategy S_k induces a mapping S_k : Statements → $[0,1]^3$. S_1 (neutrosophic): $S_1(s) := (T_1, I_1, F_1)$ in $[0,1]^3$ with no further constraint. S_2 (probabilistic): $S_2(s) := (T_2, I_2, F_2)$ in $[0,1]^3$ subject to $T_2 + I_2 + F_2 = 1$. S_3 (entropy-derived): $S_3(s) := (P_yes, H_3, P_no)$ where $P_yes + P_no := 1$ and $H_3 := -[p \log_2 p + (1 - p) \log_2(1 - p)]$ with p = P_yes. $S_2(s) : T_2 + I_2 + F_2 = 1 \Rightarrow (T, I, F) \notin \mathscr{H}$

Proposition 9.1 (Structural exclusion of hyper-truth under S_2**).** *Under Strategy 2, hyper-truth is structurally impossible: for every statement s,* $S_2(s)$ *is not in H.*

Proof. By Definition 9.4, S_2(s) satisfies $T_2 + I_2 + F_2 = 1$, while membership in H requires $T + I + F > 1$. The two conditions are mutually exclusive. ■

Proposition 9.1 explains why S_2 is the natural baseline: any non-zero hyper-truth rate observed under S_1 is a representational gain that S_2 could not produce, a structural rather than empirical contrast.

Proposition 9.2 (Non-injectivity of the scalar projection). *Let* $\pi : [0,1]^3$ *to R be the scalar projection* $\pi(T, I, F) = T + I + F$*. Then* π *is non-injective, hence the scalar sum is sufficient for hyper-truth detection but not for the discrimination of distinct epistemic regimes.*

Proof. The triples (0.5, 0.5, 0.5) and (0, 1, 0.5) both yield π = 1.5 yet differ in their first component. ■

9.3 Experimental Design

We selected five distinct linguistic phenomena to test the models' reasoning capabilities. Logical paradoxes: statements that lead to self-contradiction (this sentence is false). Epistemic ignorance: statements whose truth value is unknown in principle (the number of stars in the universe is even). Vagueness: statements with imprecise boundaries (John is 1.75 metres tall, therefore John is tall). Ethical contradictions: dilemmas where moral principles conflict (lying to save an innocent life is morally right and wrong at the same time). Future contingencies: statements about future events not yet determined (it will rain in New York tomorrow, with tomorrow anchored to 1 May 2026).

Models and parameters: the experiment involved four OpenAI models accessed via the OpenAI Chat Completions API on 30 April 2026: gpt-4o, gpt-4-turbo, gpt-3.5-turbo, and gpt-4o-mini. All calls used temperature 0.7, default top_p, no fixed seed, and a soft response-format constraint instructing the model to return only a JSON object. Each combination of (model x phenomenon x strategy) was evaluated five times in independent API calls, yielding 4 x 5 x 5 = 100 cells per strategy and a total of 300 API calls. All code, prompts, and raw data are openly released at github.com/mleyvaz/neutrosophic-llm-logic under the MIT licence.

9.4 Results

9.4.1 Descriptive Statistics

Phenomenon	**T**	**I**	**F**	**Sum**	**n**
Contingency	0.450	0.475	0.305	1.230	20
Contradiction (Ethical)	0.605	0.530	0.470	1.605	20
Ignorance (Epistemic)	0.160	0.865	0.280	1.305	20
Paradox	0.120	0.865	0.370	1.355	20

(Logical)					
Vagueness (Fuzzy)	0.562	0.345	0.242	1.150	20

Table 9.1. Descriptive statistics for neutrosophic components under Strategy 1 by phenomenon (mean across 20 evaluations per row).

9.4.2 Hyper-Truth: Breaking the Probabilistic Constraint

Across the 100 valid Strategy-1 evaluations, the empirical hyper-truth rate is 66 / 100 = 0.660. The 95 percent Wilson score confidence interval for a binomial proportion with 66 successes in 100 trials is [0.563, 0.747]. The lower bound 0.563 already exceeds any reasonable null hypothesis of zero hyper-truth, and the entire interval is well above the structural bound of zero implied by Proposition 9.1.

Phenomenon	**Hyper-truth k**	**Total n**	**Rate**
Contingency (Future)	14	20	70.0%
Contradiction (Ethical)	19	20	95.0%
Ignorance (Epistemic)	11	20	55.0%
Paradox (Logical)	10	20	50.0%
Vagueness (Fuzzy)	12	20	60.0%

Table 9.2. Hyper-truth rate by phenomenon under Strategy 1.

A Pearson chi-square test of independence between phenomenon class and hyper-truth status (5 x 2 contingency table) yields chi-square = 11.32 with df = 4 and p = 0.023, allowing rejection of independence at α = 0.05. One-vs-rest Fisher exact tests identify ethical contradiction as the only phenomenon whose hyper-truth rate is significantly higher than the rest of the dataset (odds ratio = 13.34, p = 0.0014). The chi-square result confirms that hyper-truth incidence is heterogeneous across phenomena and that ethical contradiction is the principal driver of that heterogeneity.

9.4.3 Comparison of Neutrosophic and Probabilistic Strategies

Phenomenon	S1 T	S2 T	δ T	S1 I	S2 I	δ I
Contingency	0.450	0.355	+0.095	0.475	0.470	+0.005
Contradiction (Ethical)	0.605	0.338	+0.267	0.530	0.515	+0.015
Ignorance (Epistemic)	0.160	0.231	-0.071	0.865	0.482	+0.383
Paradox (Logical)	0.120	0.000	+0.120	0.865	0.900	-0.035
Vagueness (Fuzzy)	0.562	0.450	+0.112	0.345	0.305	+0.040

Table 9.3. Strategy shifts δ T and δ I per phenomenon. Positive values indicate that the probabilistic constraint suppresses the corresponding component relative to the neutrosophic protocol.

The largest absolute strategy shifts are observed for ethical contradiction in the truth component (δ T = +0.267) and for epistemic ignorance in the indeterminacy component (δ I = +0.383). Both are positive, indicating that the probabilistic constraint of Strategy 2 suppresses precisely the components that Strategy 1 allows the model to communicate.

9.5 Cross-Vendor Replication

Mason (2026) independently replicated and extended an earlier release of this work across five additional model families from five different vendors (Anthropic, Meta, DeepSeek, Alibaba, Mistral), reporting hyper-truth in 84 percent of unconstrained evaluations. This confirms that the phenomenon is cross-vendor rather than an OpenAI-specific artefact. Chapter 5 of this book reports the subsequent multi-vendor study at 1,830 evaluations, which extended the protocol space (S4, S4-N, S4-O.A, S4-O.C) and added the auditor-effect crossover; the results confirm that hyper-truth is a structural property of the foundation-model evaluation landscape rather than an artefact of any single vendor, model, or protocol.

9.6 Discussion

Our results are consistent with the hypothesis that under unconstrained neutrosophic prompting, current language models declare hyper-truth at a non-trivial rate (66.0 percent), with the highest rate occurring for ethical contradiction (95 percent) and the chi-square test rejecting independence between phenomenon and hyper-truth at $\alpha = 0.05$.

We do not claim that hyper-truth is an intrinsic latent variable directly observed inside the model. Strategy 1 explicitly affords the model the option of returning three independent components on $[0, 1]$; the resulting frequency of hyper-truth is therefore a representational affordance finding, not a latent-variable measurement. The contribution is correspondingly framed as: unconstrained neutrosophic prompting elicits a class of declared epistemic states that probabilistic prompting cannot represent by construction. This is structural rather than empirical superiority: Strategy 2 is excluded from the hyper-truth region by construction, so any non-zero rate under Strategy 1 is a representational gain that Strategy 2 could not produce.

The relationship to other uncertainty-quantification frameworks is straightforward. Semantic entropy estimates indeterminacy from the distribution of paraphrases of the model output; it remains a probabilistic measure and therefore cannot represent hyper-truth. SelfCheckGPT performs consistency checks across stochastic samples and reports a binary or scalar consistency score, which collapses the conflict-versus-ignorance distinction we recover. Conformal abstention addresses when a model should refuse to answer; it does not describe the structure of the uncertainty when the model does answer. The neutrosophic framework is complementary to these approaches: it provides a richer descriptive language for the epistemic state, on top of which calibration and abstention policies can still operate.

9.7 Conclusion

Chapter 9 has presented an empirical investigation of neutrosophic logic applied to declared epistemic uncertainty in large language models. The unconstrained T/I/F protocol elicits hyper-truth in 66.0 percent of evaluations across the four-model ensemble, with Wilson 95 percent confidence interval [0.563, 0.747]. The highest rates were observed in ethical contradictions and future contingencies; only ethical contradiction is significantly above the pooled baseline at $\alpha = 0.05$. Mason's independent replication confirms cross-vendor generality at 84 percent across five additional vendors. Chapter 10 turns to a related but distinct question: not whether language

models can produce hyper-truth when permitted, but whether existing detection methods can recognise the simultaneous-evidence regime in which hyper-truth manifests. The answer, we will show, is structurally negative under the dominant softmax-normalised NLI architecture.

Chapter 10

Hallucination Detection Beyond Softmax: The Trichotomy

Hallucination detection methods for large language models have converged on a shared architectural commitment: deriving both a truth-support score T and a truth-contradiction score F from the same softmax-normalised natural-language inference (NLI) head. This chapter shows that this commitment introduces a previously unstated measurement-space constraint. The softmax normalisation T + neutral + F = 1 implies T + F at most 1, mathematically precluding the regime in which a model simultaneously produces substantial evidence for and against the same proposition. We prove this as a one-line corollary of the softmax simplex and argue that it provides a structural, not contingent, explanation for the empirical plateau of hallucination-detection F1 scores in the 0.65 to 0.80 range. To recover the excluded regime, we propose a dual-NLI protocol in which T and F are computed by two independently trained NLI models, and we validate it on a fifty-pair synthetic benchmark.

10.1 The Single-NLI Design Pattern

We survey three families of state-of-the-art hallucination detectors and identify the shared structural commitment. Semantic entropy (Farquhar et al., 2024) clusters multiple stochastic samples of a language model into semantically equivalent groups using a single NLI model and treats high entropy across groups as a hallucination signal. The NLI model used (Microsoft DeBERTa-v3-large-mnli) is softmax-normalised. SelfCheckGPT (Manakul, Liusie, and Gales, 2023) compares a primary response against N stochastic resamples via NLI; high inconsistency signals hallucination. The NLI variant uses RoBERTa-large-MNLI, softmax-normalised. FActScore (Min et al., 2023) decomposes responses into atomic facts and verifies each against external knowledge using NLI. The standard pipeline employs DeBERTa-v3 with softmax output. Calibration baselines (Kuhn et al., 2023; Tian et al., 2023) train auxiliary classifiers whose input features almost always include the entailment and contradiction probabilities from a softmax-normalised NLI model. In every case, T (entailment) and F (contradiction), when used at all, are drawn from the same softmax-normalised output head.

10.2 The Softmax-Paraconsistency Impossibility Theorem

Theorem 10.1 (Softmax-paraconsistency impossibility). *Let M be a softmax-normalised NLI model emitting probabilities for three classes {entailment, neutral, contradiction} for a premise-hypothesis pair (p, h). Define T(p, h) = entailment(p, h) and F(p, h) = contradiction(p, h). Then for all pairs (p, h), T(p, h) + F(p, h) at most 1, with equality if and only if neutral(p, h) = 0.*

Proof. By the definition of softmax, $p_{\text{ent}} + p_{\text{neu}} + p_{\text{con}} = 1 \Rightarrow T + F \leq 1$, with each component in $[0, 1]$. Substituting T = entailment and F = contradiction yields T + F = 1 - neutral at most 1. Equality holds when neutral = 0. ■

Corollary 10.1 (Detector impossibility). *Any hallucination detector that derives both its truth signal T and its falsity signal F from the same softmax-normalised NLI model is incapable of assigning T + F > 1 to any input. Such a detector cannot in principle distinguish responses that simultaneously support and contradict a reference proposition from any other configuration on the T + F at most 1 simplex.*

Geometric interpretation. The reachable region under softmax-normalised NLI is the lower triangle of the unit square: a set of measure one half. The excluded regime is the upper triangle, of equal measure but structurally inaccessible. No amount of additional training data, model scale, or calibration fine-tuning moves the reachable boundary. Under the dual-NLI protocol introduced in Section 10.4, T and F are computed by independent models and the full unit square becomes reachable.

10.3 Why This Explains the F1 Plateau

Tonmoy et al. (2024) report a comprehensive survey of hallucination-mitigation techniques and find that detection F1 scores cluster in the 0.65 to 0.80 range across three years of methodological development. The standard explanation invokes labelling noise or model-architectural ceilings. We propose a structural explanation. If a non-trivial fraction of language-model responses occupy the excluded regime, then any detector constrained to the lower simplex must misclassify them by construction. The misclassifications appear as labelling errors but are mathematical consequences of the measurement space. F1 cannot rise above the ceiling set by the unmeasurable subset, regardless of detector sophistication. The empirical magnitude of this effect

on frontier language models is the subject of forthcoming work; here we establish only that the geometry permits it.

10.4 The Dual-NLI Protocol

Definition 10.1 (Dual-NLI protocol). *To remove the softmax constraint while preserving the rest of the standard pipeline, compute T and F from two independently trained NLI models: T(p, h) = entailment_A(p, h) and F(p, h) = contradiction_B(p, h), where Model A is fine-tuned on MNLI (general-domain entailment) and Model B is fine-tuned on FEVER plus a scientific contradiction corpus (specialised for contradiction detection). Each model remains softmax-normalised internally, but T and F are no longer drawn from the same softmax simplex. The constraint T + F at most 1 is eliminated.*

This architecture is related to deep ensembles in that it combines outputs from independently trained models to capture distributional structure that a single model cannot represent. It differs in that we do not average predictions but extract structurally distinct signals (entailment versus contradiction) from models specialised for each, and the combination is not designed to estimate predictive variance but to populate a previously inaccessible region of the measurement space. The indeterminacy component I, when desired, is computed independently from a third channel: self-consistency disagreement across N stochastic resamples, complemented by hedging-language density and refusal-marker detection.

10.5 Synthetic Validation

To verify in a transparent and fully reproducible setting that the excluded regime is reachable under the dual-NLI protocol but not under a single-NLI softmax simulation, we constructed a corpus of 50 premise-hypothesis pairs distributed across four classes. Entailment (n = 10): high token overlap, no negation. Contradiction (n = 10): high token overlap with explicit negation, for instance Paris is the capital of France versus Paris is not the capital of France. Neutral (n = 10): low token overlap, no negation, unrelated content. Simultaneous-evidence (n = 20): high token overlap with localised negation or qualification patterns, for instance Paris is the capital of France, but it is also not the capital. These items deliberately combine an affirmative content layer with an internal-negation layer.

Class	**n**	**mean T**	**mean F**	**mean T+F**	**% above**

					diagonal
Entailment	10	0.943	0.000	0.943	0%
Contradiction	10	0.120	0.650	0.770	0%
Neutral	10	0.029	0.065	0.094	0%
Simultaneous-evidence	20	0.464	0.872	1.337	100%
Total / mean	50	0.398	0.420	0.818	40%

Table 10.1. Per-class scoring statistics on the synthetic-validation set (n = 50). Under the single-NLI softmax simulation, 0 of 50 pairs are flagged above the diagonal, exactly as Theorem 10.1 predicts.

Per-class statistics under the dual-NLI protocol show that entailment, contradiction, and neutral classes lie inside the softmax-reachable lower simplex. The simultaneous-evidence class lies entirely above the diagonal: every one of the 20 hand-crafted pairs is recovered as T + F > 1 by the dual-NLI scoring (100 percent), while none is recovered by the single-NLI simulation (0 percent), as Theorem 10.1 demands. We additionally compute the Pearson correlation between T_A and 1 - F_B across the 50 pairs as an independence diagnostic. We obtain r = 0.167, consistent with substantive independence: Model B's signal is not reducible to the complement of Model A's. The dual-NLI architecture is therefore not equivalent to a single-NLI rebrand in this constructed setting.

10.6 Discussion

The softmax constraint is so embedded in standard NLI training and inference pipelines that it operates as an unstated assumption rather than an explicit choice. None of the surveyed detection methods discusses the constraint or its implications. We hypothesise that the constraint was inherited from MNLI's framing as a three-class classification problem and propagated through the literature without re-examination. This is a familiar pattern in machine learning research: methodological constraints embedded in tooling propagate downstream until someone formalises them.

Three immediate implications follow. First, reported F1 scores in the hallucination-detection literature systematically underestimate the true detection ceiling for the unconstrained measurement space; comparisons across methods remain valid, but absolute ceilings in the 0.80 to 0.85 range are likely artefacts of the constraint, not optimal performance. Second, calibration-based methods that train auxiliary classifiers on softmax-normalised NLI features inherit the constraint and cannot be improved without protocol-level changes. Third, constitutional AI evaluation pipelines that aggregate single-NLI scores across constitutional principles compound the constraint at each principle, producing systematic blind spots in the simultaneous-evidence regime.

10.7 Limitations

Theorem 10.1 assumes both T and F are derived from softmax-normalised NLI. Some recent detectors use ensemble methods or extracted features; these do not strictly fall under the theorem and should be evaluated case by case. The synthetic validation uses heuristic scorers, not state-of-the-art NLI models. The geometric property, that simultaneous-evidence items lie above the diagonal and softmax simulation cannot reach them, is robust to the choice of scorer; the numerical magnitudes are not. The dual-NLI protocol assumes Model A and Model B are sufficiently independent; the synthetic test is one diagnostic, and full empirical evaluation against trained NLI models is required to confirm the property at scale. The rate of simultaneous-evidence behaviour in real language-model outputs is unknown at the present time, and a forthcoming empirical companion paper addresses this directly.

10.8 Conclusion

Hallucination detection in large language models has been pursued under an unstated structural constraint: when entailment T and contradiction F are derived from the same softmax-normalised NLI model, the constraint T + F at most 1 mathematically excludes the regime in which the model simultaneously supports and contradicts a proposition. We have proved this as a one-line corollary of the softmax simplex (Theorem 10.1), demonstrated in a synthetic setting that the excluded regime is non-empty, and provided a dual-NLI protocol that recovers it at moderate computational cost. The contribution is architectural. The magnitude of simultaneous-evidence behaviour in frontier language-model outputs, and the F1 lift achievable by augmenting standard

detectors with the dual-NLI feature, are empirical questions reserved for forthcoming work and synthesised in the unified auditing protocol of Chapter 11.

Chapter 11

A Unified Epistemic Auditing Protocol

Chapters 9 and 10 developed two distinct protocols for the empirical auditing of large language models: the unconstrained T/I/F neutrosophic protocol that elicits hyper-truth in Chapter 9, and the dual-NLI protocol that recovers the simultaneous-evidence regime in Chapter 10. Chapter 5 extended the first protocol to the multi-vendor setting with five elicitation strategies (S1, S4, S4-N, S4-O.A, S4-O.C) and 1,830 main evaluations. This chapter synthesises these protocols with the Type-k framework of Chapter 3 to propose a unified epistemic auditing pipeline. The pipeline takes as input a target language model, an evaluation task, and a Type-k specification (typically k = 2 in industrial applications); it returns a Type-k neutrosophic profile of the model's behaviour on the task that can be aggregated across vendors, decomposed by phenomenon, and compared against external uncertainty-quantification baselines. The protocol is designed to be executable on the OpenRouter API or any equivalent multi-vendor gateway, and the chapter closes with concrete recommendations for industrial deployment.

11.1 The Three Component Protocols

11.1.1 The S4-O Family (Chapter 5)

The S4-O family of prompts elicits neutrosophic evaluations on extended ranges. S4-O.A operates on the overset $[0,2]^3$ and uses the analogy of an employee who works overtime to motivate values greater than 1. S4-O.B operates on the underset $[-1,1]^3$ and reserves negative values for adversarial regimes in which the output actively subtracts trust. S4-O.C operates on the offset $[-1,2]^3$ and is the peer-evaluation variant in which one model audits the output of another. S4-O.D is the proposed plithogenic per-attribute offset variant, which combines the offset range with per-attribute decomposition; this variant has not yet been run at scale.

11.1.2 The Dual-NLI Protocol (Chapter 10)

The dual-NLI protocol replaces the single softmax-normalised NLI head of standard hallucination detectors with two independently trained NLI models. Model A is fine-tuned on MNLI for general-domain entailment; Model B is fine-tuned on FEVER plus a scientific contradiction corpus. The truth signal T is the entailment score from Model A; the falsity signal F is the contradiction score from Model B. The constraint T + F at most 1 is eliminated, and the full

unit square becomes reachable. An indeterminacy channel I can be computed independently from self-consistency disagreement across N stochastic resamples.

11.1.3 The Type-k Lift (Chapter 3)

The Type-k lift takes a scalar (T, I, F) evaluation produced by either of the above protocols and extends it to a Type-2 representation by additionally eliciting nine sub-component values: the (T_T, I_T, F_T) sub-triplet that describes the model's confidence in its own truth assessment, the (T_I, I_I, F_I) sub-triplet that describes its confidence in its own indeterminacy assessment, and the (T_F, I_F, F_F) sub-triplet that describes its confidence in its own falsity assessment. The lift is computationally inexpensive (it doubles or triples the prompt length and the number of API calls per evaluation), and Chapter 3 documented that 24.7 percent of responses across 2,419 evaluations require Type-2 representation under extended protocols.

11.2 The Unified Pipeline

The unified epistemic auditing pipeline proceeds in seven steps. Step 1: target specification. The auditor specifies the target language model, the evaluation task (typically a list of statements or premise-hypothesis pairs), the Type-k depth (typically k = 2), and the vendor pool (the set of additional models that will serve as peer evaluators for the auditor-effect crossover). Step 2: Type-1 elicitation. The target model is prompted under all five S4-O variants for each statement in the task, producing five Type-1 evaluations per statement. Each evaluation is checked against the admissible range of the corresponding variant; out-of-range responses are flagged for Type-2 lifting in Step 4.

Step 3: dual-NLI computation. For each (premise, hypothesis) pair in the task, the dual-NLI protocol of Chapter 10 is applied: Model A returns the entailment score, Model B returns the contradiction score, and the indeterminacy is computed from self-consistency disagreement. The resulting (T, F, I) triplet is recorded alongside the corresponding Type-1 evaluation from Step 2. Step 4: Type-2 lifting. For each Type-1 evaluation flagged in Step 2 as requiring extended representation (any component outside $[0, 1]$), the target model is re-prompted with the Type-2 elicitation template, producing the nine sub-component values. The Type-2 representation is stored in the audit log.

Step 5: peer-evaluation crossover. For each statement in the task, each model in the vendor pool is prompted under S4-O.C to audit the output of the target model. The resulting cross-vendor

matrix produces the auditor-effect findings of Chapter 5. Step 6: baseline comparison. The hyper-truth rate is computed from the S1 protocol, the Kuhn semantic entropy is computed across N stochastic resamples per statement, and the strict SelfCheckGPT score is computed using the strict equivalence judge documented in Chapter 5. The three quantities are reported alongside the Type-2 profile to enable cross-paradigm comparison. Step 7: plithogenic decomposition. The S4-N protocol is applied to a subset of statements to elicit per-attribute decomposition, and the resulting plithogenic tensor is computed following Definition 5.3.

11.3 Recommendations for Practitioners

Based on the empirical findings of Chapters 9 and 10 and the multi-vendor study of Chapter 5, we make four concrete recommendations for the industrial deployment of the unified auditing protocol. First, audit pipelines should consult at least two heterogeneous evaluators and report the union of negative-coordinate findings, not their intersection. The auditor-effect crossover documented in Chapter 5 shows that gpt-4o assigns negative coordinates in 0.0 percent of S4-O.C cells, while claude-sonnet-4 and llama-4-maverick assign them in 6 to 7 percent. A pipeline that uses only gpt-4o will systematically under-report the adversarial regime that other evaluators detect.

Second, hallucination detectors should be lifted to the dual-NLI protocol whenever F1 scores in the 0.65 to 0.80 plateau are observed. The structural ceiling identified in Theorem 10.1 means that no amount of single-NLI optimisation will exceed approximately 0.80 to 0.85 in F1; the dual-NLI lift is the only architectural remedy that addresses the underlying measurement-space limitation.

Third, for high-stakes decisions requiring maximal epistemic expressiveness, the auditing protocol should be run at Type-2 depth and the results aggregated using the recursive SVNWA operator of Definition 3.6. The 64-fold gain in structural information per case that the plithogenic tensor provides over the scalar protocol is operationally significant in safety-critical settings such as clinical decision support, criminal-justice risk assessment, and high-stakes financial advice.

Fourth, for audit and certification purposes, both the Type-1 and Type-2 aggregated scores should be reported, alongside the dual-NLI scores and the standard baselines (Kuhn semantic entropy and strict SelfCheckGPT). The redundancy serves two purposes. It quantifies the information lost in the Type-1 projection by direct comparison with the Type-2 profile; and it provides an external

cross-check on the neutrosophic findings via the established uncertainty-quantification literature, even though Chapter 5 documented that the two paradigms measure orthogonal aspects.

11.4 Open Empirical Questions

Three empirical questions remain open and define the natural continuation of the empirical work reported in this book. First, the rate of simultaneous-evidence behaviour in real frontier-language-model outputs is unknown at the time of writing. The synthetic validation of Chapter 10 establishes that the excluded regime is non-empty and reachable under the dual-NLI protocol; the magnitude of the regime on real outputs from Claude Opus 4.7, GPT-5, and Gemini 3, evaluated against TruthfulQA, HaluEval, and FActScore, is the subject of a forthcoming companion paper. [PENDING: experiment A with real frontier models on TruthfulQA, HaluEval, and FActScore.]

Second, the auditor-effect finding of Chapter 5 was instantiated with three peer auditors (gpt-4o, claude-sonnet-4, llama-4-maverick). Whether the structural isolation of gpt-4o extends to other GPT-class models or whether it is specific to the reinforcement-learning-from-human-feedback lineage of gpt-4o is open. A natural follow-up study would include at least five GPT-class auditors and at least three non-GPT auditors, with stratified analysis across architectural families.

Third, the Type-2 elicitation templates used in Chapter 3 and recommended for Step 4 of the unified pipeline have not yet been calibrated against expert-philosophical evaluations of the same statements. The five canonical statements used in Chapters 5 and 9 (paradox, ignorance, vagueness, ethical contradiction, contingency) are particularly well-suited for such a calibration exercise: they are well-studied in the philosophical literature, and a panel of five to ten expert annotators could provide ground-truth Type-2 evaluations against which the model self-reports could be compared. This calibration study is reserved for future work.

11.5 Conclusion

The unified epistemic auditing protocol introduced in this chapter combines three independent lines of work into a single executable pipeline. The S4-O family of elicitation prompts provides the Type-1 and Type-2 input data; the dual-NLI protocol provides the architecturally complete hallucination-detection back end; and the Type-k framework of Chapter 3 provides the recursive representational machinery that captures both the scalar epistemic state and the meta-uncertainty about that state. The pipeline is designed for industrial deployment, requires only standard multi-vendor API access, and produces a Type-k neutrosophic profile that is interpretable, comparable

across vendors, and aligned with the formal apparatus of Parts I through III of this book. Part V closes the book with ten open problems and a research agenda for the period 2026 through 2030.

PART V

Synthesis and Open Problems

Chapter 12

Open Problems and Research Directions

The chapters of this book have introduced Type-k Neutrosophic Sets, single-valued neutrosophic tensors, plithogenic tensors, Neutrosophic Paraconsistent Logic, and the unified epistemic auditing pipeline for large language models. Each of these contributions opens a set of research questions that we have not addressed in detail. This closing chapter collects ten such questions, organised by the part of the book from which they emerge, and outlines a coherent research agenda for the period 2026 through 2030. The agenda is informed by our broader programme on pluriversal artificial intelligence alignment, on neutrosophic statistics as the natural inferential apparatus for the empirical chapters, and on the emerging field of plithogenic evaluation for high-stakes decision support.

12.1 Ten Open Problems

Problem 1: Convergence of Type-k as k tends to infinity

Remark 3.2 noted that T, I, F may be neutrosophic sets rather than scalars, yielding an infinite-dimensional construction that subsumes all finite Type-k as projections. The formal development of this case requires defining a topology on the space of Type-k elements that makes the canonical embedding φ continuous and establishing convergence conditions for recursive aggregation under the Type-k SVNWA operator. We conjecture that the natural topology is the product topology on the countable product of unit cubes and that convergence holds under mild monotonicity assumptions, but the formal verification is open.

Problem 2: Type-k distance measures and clustering of LLMs

The empirical results of Chapter 3 documented that 24.7 percent of responses across six foundation models require Type-2 representation, with rates ranging from 19.6 percent (Alibaba) to 29.9 percent (Mistral). A natural follow-up is to define Type-k distance measures that characterise the Type-2 profile of each model and to cluster the models accordingly. This would enable principled selection of language-model-based expert sources for neutrosophic multi-criteria decision-making, based on the complexity of the decision problem and the required depth of epistemic expression.

Problem 3: Decomposition algorithms for SVN and plithogenic tensors

Theorem 4.1 established that every SVN tensor admits a Tucker decomposition componentwise, but the decomposition was presented as existence. Practical algorithms that implement the decomposition efficiently, with comparable convergence properties to the standard Higher-Order Singular Value Decomposition, are not yet available. The algorithmic question is twofold: how to enforce the SVN constraint T + I + F at most 3 during the iterative updates, and how to extend the standard Tucker algorithm to the plithogenic case where the contradiction matrix C imposes additional constraints on the attribute mode.

Problem 4: Calibrated cross-cultural panels for the contradiction function

The contradiction function c of plithogenic tensors is currently elicited by lexical overlap (Jaccard) or by aggregation of language-model outputs. A natural improvement is to build a calibrated cross-cultural panel of human annotators from at least four ethical traditions (Western liberal, Andean, ubuntu, and East Asian Confucian, for instance) and to elicit the contradiction function from their agreed judgements. The resulting panel-elicited contradiction matrix would provide a ground truth against which the lexical and language-model-based estimators could be validated.

Problem 5: Mixed-effects models for the auditor-effect finding

Chapter 5 documented that gpt-4o is structurally isolated in projecting S4-O.C audits back onto $[0, 1]$ (chi-square = 83.50, df = 2, p = 7.4 x 10^{-19}). The current analysis aggregates across vendors and phenomena. A mixed-effects logistic regression with vendor and phenomenon as random effects would clarify whether the isolation is uniform across phenomena or concentrated on specific subsets, and whether it is mediated by the audited model's behaviour. We have specified the model but have not yet fit it; this is a natural extension of the empirical work.

Problem 6: Completeness of the NPL inference system

Chapter 7 introduced the Hilbert-style inference system H-NPL for Neutrosophic Paraconsistent Logic and proved eight theorems including the failure of explosion and the conservation of ontological contradictions. The completeness of H-NPL with respect to the three-valued semantics remains open. A tableau-based proof procedure for automated reasoning would also be valuable; neither has been developed at the time of writing.

Problem 7: Integration of NPL with probabilistic frameworks for I-calibration

The threshold θ of Definition 7.3 is currently a free parameter, set to 0.5 by default. A principled calibration would integrate NPL with a probabilistic framework that estimates θ from domain-specific data: for instance, by treating I as a latent variable in a Bayesian model whose posterior is fit on a labelled dataset of ontological-versus-epistemic contradictions. The integration is conceptually straightforward but has not yet been carried out.

Problem 8: Hallucination-detection benchmarks under the dual-NLI protocol

Chapter 10 established the softmax-paraconsistency impossibility theorem and validated the dual-NLI protocol on a fifty-pair synthetic benchmark. The rate of simultaneous-evidence behaviour in real frontier-language-model outputs is unknown. Three benchmarks (TruthfulQA, HaluEval, FActScore) are natural test cases, and three frontier models (Claude Opus 4.7, GPT-5, Gemini 3) are natural subjects. The expected F1 lift relative to single-NLI baselines is an empirical question that the synthetic validation cannot answer.

Problem 9: Type-k extensions of refined neutrosophic logic

Refined neutrosophic logic, introduced by Smarandache in 2013, increases cardinality at a single level by splitting T into sub-components (T_1, ..., T_p) that remain scalars. The Type-k extension of this book increases recursion depth. The two extensions are orthogonal, as noted in Chapter 3. A Refined Type-2 Neutrosophic Set would split each of the nine Type-2 scalars into sub-components, enabling simultaneous modelling of multiple sources of indeterminacy at multiple recursion depths. The formal definition is straightforward; the question is whether the empirical evidence supports the cardinality dimension as a separate axis of refinement, distinct from the recursion-depth dimension that Chapter 3 documented.

Problem 10: Connection with neutrosophic statistics

The empirical chapters of this book report standard frequentist statistics (Pearson chi-square, McNemar tests, Wilson confidence intervals, Pearson correlations). Aslam (2019) and Smarandache and Khalid (2015) have developed neutrosophic statistics as the inferential apparatus appropriate for data whose individual observations are themselves neutrosophic. The natural step is to re-analyse the 2,419-evaluation dataset of Chapter 3 and the 1,830-evaluation dataset of Chapter 5 under neutrosophic statistics, and to compare the resulting inferences with the classical analyses. We expect the qualitative conclusions (hyper-truth prevalence, cross-

vendor consistency, auditor-effect significance) to be preserved, but the quantitative confidence intervals will widen to reflect the indeterminacy of the underlying observations.

Problem 11: N-alethic Neutrosophic Logic and Type-k N-alethic Valuations

Section 3.7.3 established that Type-k Neutrosophic Sets are isomorphic to a restricted class of n-alethic neutrosophic models in which perspectives are defined by the recursion structure rather than by external agents. The full N-alethic framework, currently in preparation, introduces a preordered perspective space (Π, <=), a perspective-specific force field $phi_p i$ governing temporal evolution of valuations, and an axiom of non-synthetic stability that admits configurations ($T_p i^t$, $I_p i^t$, $F_p i^t$) with $T_p i * F_p i > 0$ as permanent fixed points. Three open questions arise immediately from this isomorphism.

First, the compositionality question: does the type-k n-alethic valuation $V_\pi^t(A) \in [0,1]^{3^k}$ inherit a natural recursion from the Type-k structure, the perspectival structure, or both? Specifically, is $[0,1]3^k SVNWA^{(k)}$ applied perspective-by-perspective and then aggregated across perspectives, or is the aggregation order reversed, and does the order matter? The backward-compatibility result of Proposition 3.1 depends on the embedding φ mapping each scalar s to (s, 0, 1 - s); in the n-alethic setting, the embedding must additionally specify the initial force vector $phi_{pi_initial}$ of the new perspective. The choice is not canonical and affects the limiting behaviour of the transition operator.

Second, the empirical calibration question: the per-vendor Type-2 rates of Table 3.2 (19.6 percent to 29.9 percent) are interpretable as measurements of inter-perspectival conflict under the n-alethic reading. Can the force vectors phi_{pi_i} be estimated from the empirical data, and does the resulting force-field representation predict cross-vendor divergence on held-out phenomena not included in the 2,419-evaluation dataset? A natural test is whether the six-vendor force field estimated on the five epistemic phenomena of Chapter 3 generalises to the twelve ethical dilemmas of Chapter 4.

Third, the non-synthetic stability question for LLMs: the n-alethic axiom of non-synthetic stability asserts that there exist formulas A and perspectives π such that $v_p i^t$(A) is a fixed point with $T_p i * F_p i > 0$ for all time t greater than some threshold. The Absorption Problem (high and persistent I = 1 under all protocols for specific phenomena) is a candidate instance. Whether it

satisfies the formal axiom depends on whether the observed persistence is genuine temporal stability or merely an artefact of stationary experimental conditions. A longitudinal replication study across model versions would clarify the distinction.

12.2 Research Agenda 2026-2030

The ten open problems above structure a coherent research agenda for the period 2026 through 2030. We organise the agenda in three phases that parallel the structure of the book.

12.2.1 Phase 1 (2026-2027): Foundations and Empirical Calibration

The first phase consolidates the formal foundations and calibrates the empirical findings against external ground truths. The priorities are Problem 1 (convergence of Type-k as k tends to infinity), Problem 3 (decomposition algorithms for SVN and plithogenic tensors), Problem 4 (calibrated cross-cultural panels for the contradiction function), and Problem 6 (completeness of H-NPL). The expected deliverables are a peer-reviewed paper establishing the infinite-dimensional Type-k construction, an open-source library implementing the Tucker decomposition for SVN tensors, a panel-elicited contradiction matrix for the fourteen ethical attributes of Chapter 4, and a completeness theorem for H-NPL with respect to the three-valued semantics.

12.2.2 Phase 2 (2027-2028): Industrial Validation and Decision Support

The second phase validates the framework on real industrial multi-criteria decision-making problems and extends the unified auditing pipeline to production deployment. The priorities are Problem 2 (Type-k distance measures for clustering), Problem 5 (mixed-effects models for the auditor-effect finding), Problem 7 (NPL integration with probabilistic frameworks), and Problem 10 (neutrosophic statistics). The expected deliverables are a deployed clinical decision support audit based on the Chapter 6 framework, a stratified analysis of the auditor-effect crossover across at least eight vendors, a Bayesian calibration of the NPL threshold θ on a labelled dataset of contradictions, and a re-analysis of the empirical datasets under neutrosophic statistics.

12.2.3 Phase 3 (2028-2030): Pluriversal Alignment and Public Audit

The third phase extends the framework to pluriversal artificial intelligence alignment and to the public auditing of frontier-language-model deployments. The priorities are Problem 8 (hallucination-detection benchmarks under the dual-NLI protocol) and Problem 9 (Refined Type-

k extensions), together with a research line on constitutional AI evaluation that combines the unified auditing pipeline of Chapter 11 with the Pol.is-based public-deliberation methodology of the Collective Constitutional AI literature. The expected deliverables are large-scale dual-NLI benchmarks on the three standard hallucination-detection corpora, a Refined Type-k formal framework, and a public audit of at least three frontier-model deployments under the unified pipeline.

12.3 Connections to Pluriversal Alignment

The Type-k framework and the NPL inferential system together provide what we consider the natural formal infrastructure for pluriversal artificial intelligence alignment: the project of building AI systems that can simultaneously accommodate value frameworks from multiple, structurally incompatible ontologies (liberal individualist, communal Andean, ubuntu, Confucian, and others that the present literature has yet to articulate). The classical alignment literature treats value pluralism as an aggregation problem: given multiple stakeholder preferences, find the social choice that maximises a defined welfare function. The NPL framework reframes the problem as a representation problem: design inference systems that can hold incompatible values simultaneously, classify their tensions as ontological or epistemic, and reason productively about the resulting configurations without forcing premature resolution.

The empirical chapters of Part IV suggest that contemporary foundation models already exhibit, in their declared epistemic states, a richer structure than the standard alignment metrics can capture. The 95 percent hyper-truth rate on ethical contradictions documented in Chapter 9, the cross-vendor consistency of the auditor-effect crossover documented in Chapter 5, and the structural centrality of human autonomy in the contradiction graph documented in Chapter 4 are all measurements that the standard probability-based alignment literature does not produce. We interpret them as preliminary evidence that the language models, when permitted, are willing to operate in regimes that classical alignment metrics treat as failure modes. The research agenda of the next four years aims to make this provisional evidence rigorous, reproducible, and actionable.

12.4 Conclusion

The book has presented an integrated formal and empirical framework for the epistemic auditing of artificial intelligence systems. Part I introduced Type-k Neutrosophic Sets, the recursive triadic structure that generalises the classical neutrosophic framework to nested uncertainty. Part II

developed the tensorial machinery (SVN tensors, plithogenic tensors) and the decision-theoretic consequences (Decision Optimality, Contradiction Visibility). Part III introduced Neutrosophic Paraconsistent Logic and connected it to the Type-k framework. Part IV applied the framework to the empirical auditing of foundation models through three complementary protocols (hyper-truth elicitation, dual-NLI hallucination detection, unified pipeline). Part V has surveyed the ten open problems that the framework opens and outlined a four-year research agenda to address them.

The central thesis of the book is that the epistemic structure of contemporary artificial intelligence systems is irreducibly triadic and recursive. The classical neutrosophic framework provides the triadic apparatus; the Type-k extension provides the recursive apparatus; the tensorial extensions provide the computational machinery; the paraconsistent extensions provide the inferential machinery; and the empirical chapters demonstrate that all of this apparatus is operationally necessary for the auditing of contemporary foundation models. The book is offered as both a self-contained reference for researchers entering the field and a working manual for industrial practitioners building auditing pipelines on top of the multi-vendor language-model infrastructure that now defines the engineering frontier of artificial intelligence.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications, 7(1), 39–59.

Abdel-Basset, M., Manogaran, G., Gamal, A., & Smarandache, F. (2019). A group decision making framework based on neutrosophic TOPSIS approach for smart medical device selection. Journal of Medical Systems, 43(2), Article 38.

Abdel-Basset, M., Mohamed, R., Zaied, A. E.-N. H., & Smarandache, F. (2019). A hybrid plithogenic decision-making approach with quality function deployment for selecting supply chain sustainability metrics. Symmetry, 11(7), Article 903.

Anderson, A. R., & Belnap, N. D. (1975). Entailment: The logic of relevance and necessity (Vol. 1). Princeton University Press.

Angelopoulos, A. N., & Bates, S. (2023). Conformal prediction: A gentle introduction. Foundations and Trends in Machine Learning, 16(4), 494–591.

Aslam, M. (2018). A new sampling plan using neutrosophic process loss consideration. Symmetry, 10(5), Article 132.

Aslam, M. (2019). A variable acceptance sampling plan under neutrosophic statistical interval method. Symmetry, 11(1), Article 114.

Atanassov, K. T. (1986). Intuitionistic fuzzy sets. Fuzzy Sets and Systems, 20(1), 87–96.

Badreddine, S., Garcez, A. d., Serafini, L., & Spranger, M. (2022). Logic Tensor Networks. Artificial Intelligence, 303, Article 103649.

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., ... Kaplan, J. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv:2212.08073.

Baltag, A., Moss, L. S., & Solecki, S. (1998). The logic of public announcements, common knowledge, and private suspicions. In Proceedings of TARK 1998 (pp. 43–56).

Beall, J. (2022). Review of Zach Weber, Paradoxes and inconsistent mathematics. Notre Dame Philosophical Reviews.

Belnap, N. D. (1977). A useful four-valued logic. In J. M. Dunn & G. Epstein (Eds.), Modern uses of multiple-valued logic (pp. 5–37). Reidel.

Bilal, M., Li, C., Alzahrani, A. K., & Aljahdali, A. K. (2025). Evaluating pancreatic cancer treatment strategies using a novel polytopic fuzzy tensor approach. Bioengineering, 13(1), Article 2.

Biswas, P., Pramanik, S., & Giri, B. C. (2016). TOPSIS method for multi-attribute group decision-making under single-valued neutrosophic environment. Neural Computing and Applications, 27(3), 727–737.

Bohr, N. (1928). The quantum postulate and the recent development of atomic theory. Nature, 121, 580–590.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

Carnielli, W., & Coniglio, M. E. (2016). Paraconsistent logic: Consistency, contradiction and negation. Springer.

Chen, L. (2020). Decomposition theorem of intuitionistic fuzzy tensors. Computational and Applied Mathematics, 39, Article 18.

Chen, L., & Chen, Z. (2019). Decomposition theorem of fuzzy tensors and its applications. Journal of Intelligent & Fuzzy Systems, 36(1), 575–581.

da Costa, N. C. A. (1963). Sistemas formais inconsistentes. Tese de Livre-Docência, Universidade Federal do Paraná.

da Costa, N. C. A. (1974). On the theory of inconsistent formal systems. Notre Dame Journal of Formal Logic, 15(4), 497–510.

da Costa, N. C. A., & Subrahmanian, V. S. (1989). Paraconsistent logics as a formalism for reasoning about inconsistent knowledge bases. Artificial Intelligence in Medicine, 1(4), 167–174.

Dempster, A. P. (1968). A generalization of Bayesian inference. Journal of the Royal Statistical Society: Series B, 30(2), 205–247.

Domingos, P. (2025). Tensor logic: The language of AI. arXiv:2510.12269.

Escobar, A. (2018). Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds. Duke University Press.

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. Nature, 630, 625–630.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People: An ethical framework for a good AI society. Minds and Machines, 28, 689–707.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning (pp. 1050–1059).

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In Proceedings of the International Conference on Machine Learning (pp. 1321–1330).

He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with disentangled attention. International Conference on Learning Representations.

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. Zeitschrift für Physik, 43, 172–198.

Hendrycks, D., Burns, C., Basart, S., Critch, A., Li, J., Song, D., & Steinhardt, J. (2021). Aligning AI with shared human values. International Conference on Learning Representations.

Hintikka, J. (1962). Knowledge and belief: An introduction to the logic of the two notions. Cornell University Press.

Huang, S., Siddarth, D., Lovitt, L., Liao, T. I., Durmus, E., Tamkin, A., & Ganguli, D. (2024). Collective Constitutional AI: Aligning a language model with public input. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency. arXiv:2406.07814.

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Machine Learning, 110(3), 457–506.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.

Jiao, J., Afroogh, S., Murali, A., Chen, K., Atkinson, D., & Dhurandhar, A. (2025). LLM ethics benchmark: A three-dimensional assessment system for evaluating moral reasoning in large language models. Scientific Reports, 15, Article 34642.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399.

Kadavath, S., Conerly, T., Askell, A., Henighan, T., Drain, D., Perez, E., Schiefer, N., Dodds, Z., DasSarma, N., Tran-Johnson, E., Johnston, S., El-Showk, S., Jones, A., Elhage, N., Hume, T., Chen, A., Bai, Y., Bowman, S., Fort, S., ... Kaplan, J. (2022). Language models (mostly) know what they know. arXiv:2207.05221.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. SIAM Review, 51(3), 455–500.

Kuhn, L., Gal, Y., & Farquhar, S. (2023). Semantic uncertainty: Linguistic invariances for uncertainty estimation in natural language generation. International Conference on Learning Representations.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems.

Leyva-Vázquez, M. Y. (2024a). Dynamic epistemic neutrosophic logic for LLM reasoning chains: NCC and IR metrics. Manuscript in preparation.

Leyva-Vázquez, M. Y. (2024b). Neutrosophic analysis of competing hypotheses with Admiralty weighting. Manuscript in preparation.

Leyva-Vázquez, M. Y. (2024c). NeutroXAI: Explainability under neutrosophic logic with component-wise attribution. Manuscript in preparation.

Leyva-Vázquez, M. Y., & Smarandache, F. (2018). Neutrosofía: Nuevos avances en el tratamiento de la incertidumbre. Pons Publishing House.

Leyva-Vázquez, M. Y., & Smarandache, F. (2026). Breaking the chains of probability: Neutrosophic logic as a new framework for epistemic uncertainty in large language models. arXiv:2605.24053.

Li, J., Cheng, X., Zhao, W. X., Nie, J.-Y., & Wen, J.-R. (2023). HaluEval: A large-scale hallucination evaluation benchmark for large language models. Proceedings of EMNLP, 6449–6464.

Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., Zhang, Y., Narayanan, D., Wu, Y., Kumar, A., Newman, B., Yuan, B., Yan, B., Zhang, C., Cosgrove, C., Manning, C. D., Ré, C., Acosta-Navas, D., Hudson, D. A., ... Hashimoto, T. (2023). Holistic evaluation of language models. Transactions on Machine Learning Research.

Lin, S., Hilton, J., & Evans, O. (2022). TruthfulQA: Measuring how models mimic human falsehoods. Proceedings of ACL, 3214–3252.

Manakul, P., Liusie, A., & Gales, M. J. F. (2023). SelfCheckGPT: Zero-resource black-box hallucination detection for generative large language models. Proceedings of EMNLP.

Mason, T. (2026). From scalars to tensors: Declared losses recover epistemic distinctions that neutrosophic scalars cannot express. arXiv:2604.09602.

Mendel, J. M., & John, R. I. (2002). Type-2 fuzzy sets made simple. IEEE Transactions on Fuzzy Systems, 10(2), 117–127.

Min, S., Krishna, K., Lyu, X., Lewis, M., Yih, W.-t., Koh, P. W., Iyyer, M., Zettlemoyer, L., & Hajishirzi, H. (2023). FActScore: Fine-grained atomic evaluation of factual precision in long-form text generation. Proceedings of EMNLP, 12076–12100.

Novikov, A., Podoprikhin, D., Osokin, A., & Vetrov, D. (2015). Tensorizing neural networks. Advances in Neural Information Processing Systems.

Oseledets, I. V. (2011). Tensor-train decomposition. SIAM Journal on Scientific Computing, 33(5), 2295–2317.

Pawlak, Z. (1982). Rough sets. International Journal of Computer and Information Sciences, 11(5), 341–356.

Priest, G. (1979). The logic of paradox. Journal of Philosophical Logic, 8(1), 219–241.

Priest, G. (1987). In contradiction: A study of the transconsistent. Martinus Nijhoff.

Priest, G. (2006). Doubt truth to be a liar. Oxford University Press.

Rivieccio, U. (2008). Neutrosophic logics: Prospects and problems. Fuzzy Sets and Systems, 159(14), 1860–1868.

Shah, S., & Zadrozny, W. (2026). Implementing tensor logic: Unifying Datalog and neural reasoning via tensor contraction. arXiv:2601.17188.

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423.

Shorinwa, O., Mei, Z., Lidard, J., Ren, A., & Majumdar, A. (2024). A survey on uncertainty quantification of large language models. arXiv:2412.05563.

Smarandache, F. (1998). Neutrosophy: Neutrosophic probability, set, and logic. American Research Press.

Smarandache, F. (2002). Neutrosophic logic: A generalization of intuitionistic fuzzy logic. Multiple-Valued Logic, 8(3), 385–438.

Smarandache, F. (2005). A unifying field in logics: Neutrosophic logic (4th ed.). American Research Press.

Smarandache, F. (2013). n-valued refined neutrosophic logic and its applications to physics. Progress in Physics, 4, 143–146.

Smarandache, F. (2016). Neutrosophic overset, neutrosophic underset, and neutrosophic offset. Pons Editions.

Smarandache, F. (2018). Plithogeny, plithogenic set, logic, probability, and statistics. Pons Publishing House.

Smarandache, F. (2023). Introduction to symbolic plithogenic algebraic structures (revisited). Neutrosophic Sets and Systems, 53.

Smarandache, F., & Khalid, H. E. (2015). Neutrosophic statistics. Pons Editions.

Tian, K., Mitchell, E., Yao, H., Manning, C. D., & Finn, C. (2023). Fine-tuning language models for factuality. arXiv:2311.08401.

Tonmoy, S. M. T. I., Zaman, S. M. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). A comprehensive survey of hallucination mitigation techniques in large language models. arXiv:2401.01313.

Valdenegro-Toro, M. (2022). A deeper look into aleatoric and epistemic uncertainty estimation. arXiv:2204.09308.

Veličković, P., Perivolaropoulos, C., Barbero, F., & Pascanu, R. (2024). Softmax is not enough (for sharp out-of-distribution). arXiv:2410.01104.

Wang, H., Smarandache, F., Zhang, Y. Q., & Sunderraman, R. (2010). Single valued neutrosophic sets. Multispace and Multistructure, 4, 410–413

Weber, Z. (2022). Paradoxes and inconsistent mathematics. Cambridge University Press.

Xiong, M., Hu, Z., Lu, X., Li, Y., Fu, J., He, J., & Hooi, B. (2024). Can LLMs express their uncertainty? An empirical evaluation of confidence elicitation in LLMs. International Conference on Learning Representations. arXiv:2306.13063.

Yadkori, Y. A., Kuzborskij, I., Stutz, D., Bachem, O., & Hennig, P. (2024). Mitigating LLM hallucinations via conformal abstention. arXiv:2405.01563.

Ye, J. (2014). A multicriteria decision-making method using aggregation operators for simplified neutrosophic sets. Journal of Intelligent & Fuzzy Systems, 26(5), 2459–2466.

Zadeh, A., Chen, M., Poria, S., Cambria, E., & Morency, L.-P. (2017). Tensor fusion network for multimodal sentiment analysis. Proceedings of EMNLP.

Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338–353.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning, Part I. Information Sciences, 8(3), 199–249.

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. Advances in Neural Information Processing Systems, 36.

TYPE-k NEUTROSOPHIC SETS

Type-k Neutrosophic Sets develops a recursive extension of neutrosophic logic: each truth, indeterminacy, and falsity component can itself be represented by a full neutrosophic triplet. The result is a hierarchy for modelling nested uncertainty, tensorial decision structures, paraconsistent reasoning, and large-language-model auditing.

- Formal Type-k hierarchy and canonical embedding
- Recursive SVNWA aggregation with Type-1 compatibility
- SVN and plithogenic tensor extensions
- Neutrosophic paraconsistent logic and Type-k alignment
- Auditing protocols for epistemic behaviour in LLMs

> *A framework for decisions where uncertainty is not noise to remove, but structure to represent.*

Florentin Smarandache | Maikel Leyva-Vazquez

NSIA PUBLISHING
ADVANCING KNOWLEDGE

ISBN 978-197250222-8
9 781972 502228

Mathematics | Decision Science | AI Evaluation